



Diogo Filipe Pinto Oliveira

Licenciado em Ciências da Engenharia Eletrotécnica e de Computadores

Deteção de Ocupação usando o Consumo de Energia

Dissertação para obtenção do Grau de Mestre em
Engenharia Eletrotécnica e de Computadores

Orientador: João Francisco Alves Martins, Professor Doutor,
Universidade Nova de Lisboa

Júri

Presidente: Nuno Filipe Silva Veríssimo Paulino, Professor Doutor, FCT/UNL
Arguente: Pedro Miguel Ribeiro Pereira, Professor Doutor, FCT/UNL



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Julho, 2020

Deteção de Ocupação usando o Consumo de Energia

Copyright © Diogo Filipe Pinto Oliveira, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

À minha família, amigos e colegas de faculdade.

AGRADECIMENTOS

Em primeiro lugar gostaria de agradecer ao meu orientador, Professor Doutor João Martins, por toda a simpatia, motivação, disponibilidade e apoio científico dado durante a elaboração deste trabalho.

Quero agradecer à minha família, nomeadamente aos meus pais e avós por todo o apoio económico e moral ao longo destes anos. Ao meu tio Sandro pelas oportunidades de trabalho durante estes últimos anos, pois sem esta oportunidade não teria chegado até aqui. Ao meu primo Gonçalo pela ajuda e compreensão enquanto trabalhamos juntos. Tenho de agradecer ao meu primo e Filipe pelo exemplo, motivação e inspiração.

Aos meus colegas e amigos de curso e faculdade por toda a ajuda, companheirismo e momentos únicos vividos, foram eles, o Pedro Morgado, André Fidalgo, Gonçalo Mendonça, Pedro Cunha, Filipe Oliveira, João Gião, João Santos, Daniel Batista, Ricardo Madeira, João Estevinha, Bernardo Silva, entre outros. Um grande obrigado aos meus amigos conterrâneos, ao caçador Filipe Alves e ao Doutor Pedro Russo pela amizade e motivação durante os últimos tempos.

Por fim, um agradecimento especial ao meu tio Manuel, por teres sido a minha estrela polar durante esta caminhada. Meu querido e amigo Tio Manel, nunca esquecerei os teus conselhos, onde quer que estejas, sei que estás orgulhoso de mim. Dedico-te esta dissertação, com muito carinho e saudade, obrigado e até sempre.

RESUMO

Segundo os dados mais recentes, em Portugal os edifícios (e.g. habitações e serviços) representam 30% do consumo final de energia, constituindo um dos sectores que mais contribui na emissão de gases com efeito de estufa. Deste modo, é fundamental tomar medidas com vista a reduzir o consumo de energia. Nos últimos anos, pesquisas recentes têm mostrado que os contadores inteligentes permitem extrair e tratar informação útil a partir do consumo de energia de edifícios e habitações, permitindo tomar ações de modo a controlar o desperdício de energia. Deste modo, a União Europeia decretou regras e medidas para a instalação destes dispositivos, com o objetivo de promover a eficiência energética neste sector.

Atualmente, com a crescente instalação de contadores inteligentes é possível estudar características intrínsecas ao perfil do consumo de energia diretamente relacionadas à utilização dos equipamentos elétricos. Esta informação torna a monitorização de ocupação nos edifícios, até aqui efetuada maioritariamente por sensores, mais eficaz e não-intrusiva, aportando menos custos de utilização e manutenção para os utilizadores.

Nesta dissertação é estudada e discutida a deteção de ocupação utilizando o consumo de energia disponibilizado pelos contadores inteligentes. Para tal é implementado um sistema de deteção de ocupação através de um algoritmo de aprendizagem supervisionada. De modo a construir o modelo de aprendizagem são utilizados o consumo agregado e desagregado de energia fornecido por contadores inteligentes e o número real de ocupantes.

A avaliação prática deste sistema de deteção de ocupação é realizada mediante vários cenários onde é estudada a influencia do mês do ano, do tipo e número de equipamentos elétricos, assim como o número de ocupantes.

Palavras-chave: Eficiência Energética, Contador Inteligente, Aprendizagem Supervisionada, Deteção de Ocupação.

ABSTRACT

According to the most recent data, buildings in Portugal (e.g. housing and services) represent 30% of final energy consumption, representing one of the sectors that most contributes to the emission of greenhouse gases. Therefore, it is essential to take measures to reduce energy consumption. In recent years, new researches have shown that smart meters allow to extract and process useful information from the energy consumption of buildings and homes, allowing to take actions to control energy waste. In this way, the European Union has enacted rules and measures for the installation of these devices, with the aim of promoting energy efficiency in this sector.

Currently, with the increasing installation of smart meters, it's possible to study the intrinsic characteristics of the energy consumption profile directly related to the use of electrical equipment. This information makes the monitoring in buildings, carried out until today mostly by sensors, more effective and non-intrusive, contributing less costs of use and maintenance for users.

In this dissertation, occupation detection is studied and discussed using the energy consumption provided by smart meters. To this end, an occupancy detection system is implemented using a supervised learning algorithm. In order to build the learning model, the aggregated and disaggregated energy consumption provided by smart meters and the actual number of occupants (i.e. ground truth) are used.

The practical assessment of this occupancy detection system is carried out using various scenarios such as: the influence of the month of the year, the type and number of electrical equipment, as well as the number of occupants.

Keywords: Energy Efficiency, Smart Meter, Occupancy Detection, Supervised Learning, Occupancy Detection.

ÍNDICE

Lista de Figuras	xv
Lista de Tabelas	xvii
Siglas	xix
1 Introdução	1
1.1 Enquadramento e Motivação	2
1.2 Objetivos	4
1.3 Estrutura da Dissertação	4
2 Estado de Arte	7
2.1 Detecção de ocupação em edifícios	8
2.2 Detecção de pessoas através do consumo de energia	9
2.3 Conhecimento prévio	16
2.3.1 Conceitos Fundamentais	16
2.3.2 Aprendizagem Supervisionada	16
2.3.3 K-nearest neighbor	16
2.3.4 Avaliação	17
3 Arquitetura do Sistema	21
3.1 Modelo Conceptual	22
3.1.1 Modelo Arquitetural	22
3.2 Implementação	25
3.2.1 Tecnologias Utilizadas	25
3.2.2 Modelo de Consumo de Energia	25
3.2.3 Sistema de Gestão de Dados	27
3.2.4 Sistema de Detecção de Ocupação	28
3.2.5 Visualização de Ocupação	31
4 Testes e Resultados	35
4.1 Cenários Experimentais	36
4.2 Cenário 1	36
4.3 Cenário 2	41

ÍNDICE

4.4	Cenário 3	45
4.5	Cenário 4	49
4.6	Análise dos Resultados	53
5	Conclusão	55
5.1	Conclusões Finais	55
5.2	Trabalhos Futuros	56
	Bibliografia	57

LISTA DE FIGURAS

1.1	Consumo de energia elétrica em Portugal por setor (GWh). Adaptado de [2].	2
1.2	Percentagem de equipamentos elétricos em edifícios residenciais ao longo os anos. [3].	3
2.1	Algoritmo de deteção de ocupação baseado num método utilizando um valor limiar. Adaptado de [17].	12
2.2	Precisão dos algoritmos variando o tamanho do conjunto de treino no intervalo (1, 5, 10, 15, 20, 30)%. Adaptado de [18].	14
2.3	Framework SHARK para deteção de ocupação não intrusiva sem fase de treino. Adaptado de [28].	15
2.4	Pseudocódigo do algoritmo KNN	17
3.1	Modelo Arquitetural do Sistema	22
3.2	Arquitetura do Sistema de Gestão de Dados	23
3.3	Arquitetura do Sistema de Gestão de Dados	24
3.4	Arquitetura da camada Ocupação	24
3.5	Arquitetura do modelo de consumo de energia [34].	26
3.6	Procedimento para gerar os dados de consumo de energia e ocupação	28
3.7	Módulo de aprendizagem do modelo K-Nearest Neighbor	29
3.8	Exatidão dos modelos variando o valor de K	30
3.9	Módulo de classificação do modelo K-Nearest Neighbor	31
3.10	Módulo de avaliação do modelo K-Nearest Neighbor	31
3.11	Resultado da aplicação do algoritmo antes de aplicar a média móvel	32
3.12	Módulo da média móvel do nº de ocupantes	32
3.13	Resultado da aplicação do algoritmo após aplicar a média móvel	33
4.1	Diagrama de caixa referente à exatidão do MDO	38
4.2	Ocupação Real e Ocupação Detetada - Cenário 1.1.1	40
4.3	Curva de carga e ocupação real	41
4.4	Diagrama de caixa referente à exatidão do MDO	42
4.5	Ocupação Real e Ocupação Detetada - Cenário 2.3	44
4.6	Curva de carga e ocupação real	45
4.7	Diagrama de caixa referente à exatidão do MDO	46

LISTA DE FIGURAS

4.8	Ocupação Real e Ocupação Detetada - Cenário 3.1.1	48
4.9	Curva de Carga e Ocupação Real	49
4.10	Diagrama de caixa referente à exatidão do MDO	50
4.11	Ocupação Real e Detetada - Cenário 4.5	52
4.12	Curva de Carga e Ocupação Real	53

LISTA DE TABELAS

2.1	Matriz de Confusão	17
3.1	Equipamentos elétricos e sua relação com a ocupação.	27
4.1	Exatidão do MA, MC e MAV utilizando a energia desagregada - Cenário 1 . .	37
4.2	Relatório de Classificação	39
4.3	Matriz de Confusão	39
4.4	Exatidão do MA, MC e MAV utilizando a energia desagregada - Cenário 2 . .	42
4.5	Relatório de Classificação	43
4.6	Matriz de Confusão	43
4.7	Exatidão do MA, MC e MAV utilizando a energia desagregada - Cenário 3 . .	45
4.8	Relatório de Classificação	47
4.9	Matriz de Confusão	47
4.10	Exatidão do MA, MC e MAV utilizando a energia desagregada - Cenário 4 . .	50
4.11	Relatório de Classificação	51
4.12	Matriz de Confusão	51

SIGLAS

AVAC Aquecimento, Ventilação e Ar Condicionado.

BD Base de Dados.

BL Base Learning.

ECO Electricity Consumption and Occupancy.

EDP Energias de Portugal.

FN Falso Negativo.

FP Falso Positivo.

GPS Global Positioning System.

HMM Hidden Markov Model.

Hz Hertz.

KNN K-Nearest Neighbor.

MA Módulo de Aprendizagem.

MAV Módulo de Avaliação e Visualização.

MC Módulo de Classificação.

MCE Modelo de Consumo de Energia.

MDO Modelo de Detecção de Ocupação.

MIT Multiview-based Iteration.

ML Machine Learning.

MTD Módulo de Tratamento de Dados.

NILM Non-intrusive Appliance Load Monitoring.

NILM Non-intrusive Monitoring.

NIOM Non-intrusive Occupancy Monitoring.

NL Non-intrusive Learning.

PCA	Principal Component Analysis.
SDO	Sistema de Detecção de Ocupação.
SFS	Sequencial Foward Selection.
SGD	Sistema de Gestão de Dados.
SHARK	Sparce Human Action Recovery Detection Knowledge.
SVM	Support Vector Machine.
THR	Thresholding.
TL	Transfer Learning.
UE	União Europeia.
VN	Verdadeiro Negativo.
VP	Verdadeiro Positivo.
W	Watt.
Wh	Watt-hora.

INTRODUÇÃO

Este primeiro capítulo pretende dar uma perspetiva geral do trabalho, apresentando o enquadramento e motivação desta dissertação. Posteriormente são apresentados os objetivos traçados para esta dissertação que suportam o desenvolvimento de uma metodologia de deteção de ocupação não intrusiva através do consumo de energia elétrica, normalmente disponibilizada por contadores inteligentes instalados nas habitações. Em último lugar, apresenta-se a estrutura do documento, discriminando os pontos fundamentais de cada capítulo.

1.1 Enquadramento e Motivação

De acordo com os dados mais recentes, disponibilizados pela Direção Geral da Energia e Geologia, o setor dos edifícios em Portugal é responsável pelo consumo de aproximadamente 30% da energia final [1]. Na Figura 1.1 é possível verificar a contribuição de cada setor para o gasto energético total, sendo o setor da indústria aquele que maior peso tem no consumo total, seguido dos edifícios domésticos e não domésticos, com menos peso aparecem os edifícios do estado e o setor da agricultura. Apesar disso, segundo dados da Direção Geral da Geologia e Energia [1], através da implementação de medidas com efeito na eficiência energética mais de 50% deste consumo pode ser reduzido. Esta redução, equivalente a 400 milhões de toneladas de dióxido de carbono cumpre, quase na totalidade, o compromisso da UE no âmbito do Protocolo de Quioto.

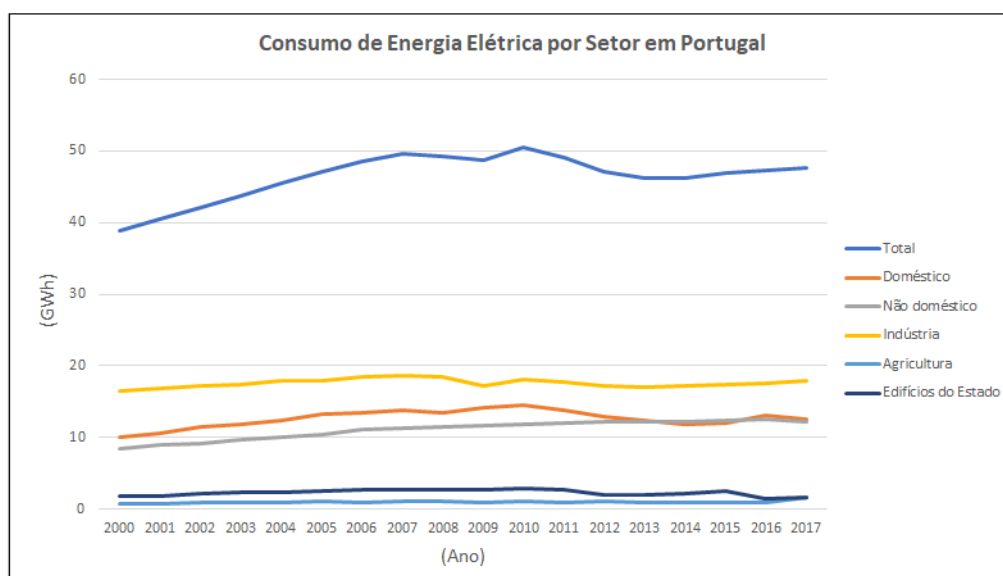


Figura 1.1: Consumo de energia elétrica em Portugal por setor (GWh). Adaptado de [2].

Assim sendo, é importante pesquisar e implementar medidas que melhorem a eficiência energética dos edifícios. Uma forma simples de reduzir o desperdício de energia e aumentar a eficiência energética passa por conhecer o nível de ocupação dos edifícios, sendo possível tomar ações de modo a regular o funcionamento dos equipamentos elétricos. Na Figura 1.2 é possível verificar a evolução da percentagem de equipamentos usados nas habitações em Portugal, onde se pode constatar o aumento da utilização de sistemas de aquecimento, ventilação e ar condicionado.

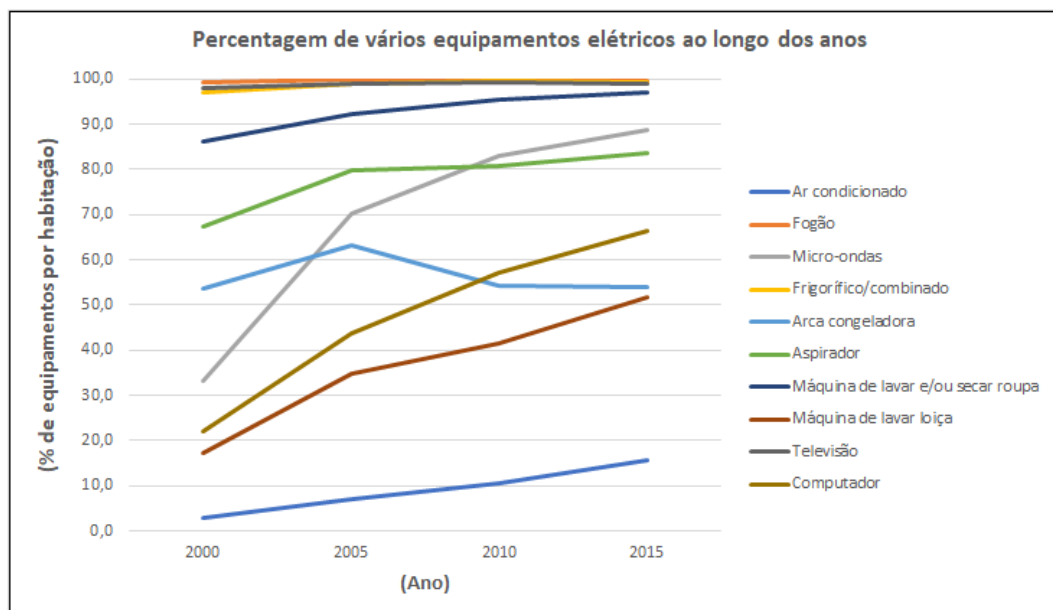


Figura 1.2: Percentagem de equipamentos elétricos em edifícios residenciais ao longo os anos. [3].

Pesquisas recentes mostraram que a gestão de energia em edifícios e habitações pode ser otimizada utilizando a informação de ocupação, nomeadamente no controlo de sistemas AVAC [4–6]. Deste modo, é possível poupar energia automaticamente nos períodos de desocupação i) desligando dispositivos (i.e., cargas elétricas) inativos, como televisões, consolas de jogos ou box de tv e Internet; ii) controlando equipamentos relacionados ao conforto interior como sistemas de aquecimento, ventilação e ar condicionado, aquecedores ou desumidificadores, operando-os remotamente; iii) desligando ou ajustando a luminosidade para reduzir o desperdício de energia; iv) mudando o estado de stand-by dos computadores e outros equipamentos de IT para níveis de consumo inferiores [7].

Nos últimos anos, a instalação de contadores inteligentes tem-se tornado frequente. De acordo com a Diretiva 2009/72/CE (eletricidade) do Parlamento Europeu e do Conselho da União Europeia, para promover e melhorar a eficiência energética, os Estados-Membros estão obrigados a avaliar a instalação de sistemas de contadores inteligentes de eletricidade. A diretiva estabelece que até 2020, 80% dos contadores de energia sejam inteligentes [8]. Como tal, as empresas devem inovar os serviços de gestão de energia, aplicando tarifários energéticos progressistas e, introduzindo contadores e redes inteligentes [9].

Atualmente, cerca de 45 milhões de contadores inteligentes foram instalados em Itália, Suécia e Finlândia ao passo que, na Alemanha, a instalação destes equipamentos é obrigatória em edifícios novos e renovados [10]. De acordo com dados da Comissão Europeia, cerca de 200 milhões de contadores inteligentes devem ser instalados até 2020 [11].

Em Portugal, a Edp Distribuição tem vindo a instalar contadores inteligentes sendo

expectável que até 2020, em Portugal, 60% dos utilizadores disponham desta nova tecnologia. A transição para estes novos dispositivos de gestão de energia requer uma adaptação por parte do utilizador comum, com vista a aproveitar todas as vantagens e funcionalidades destes contadores inteligentes.

Estes dispositivos permitem aos utilizadores melhorar a eficiência energética das suas casas, monitorizando e analisando diariamente o padrão do seu consumo, simulando ciclos horários e ainda programando alertas automáticos caso o consumo ultrapasse um determinado limite [12]. A grande vantagem da utilização dos contadores inteligentes passa pelo maior controlo do consumo de energia, permitindo também outras funcionalidades, como por exemplo: efetuar leituras automáticas, emissão de faturas sem estimativas; realizar operações remotas (e.g. alterar a tarifa de faturação ou efetuar despiste de avarias).

Apesar do principal objetivo da instalação de contadores inteligentes seja melhorar os processos de leitura e faturação, assim como efetuar operações remotas, estes dispositivos permitem às companhias de produção e distribuição de energia obter mais informações através do perfil do consumo energético dos seus clientes. Posteriormente e caso o utilizador comum permita, essa informação pode ser utilizada em outras aplicações, como na monitorização de ocupação, identificando as características na curva de carga da instalação que melhor se relacionam com o nível de ocupação. Assim sendo, a principal motivação desta dissertação passa por desenvolver um método de deteção de ocupação menos intrusivo que outros métodos baseados na utilização de sensores.

1.2 Objetivos

O principal objetivo desta dissertação passa por desenvolver um modelo de monitorização de ocupação através do consumo de energia registado por um contador de energia inteligente. O modelo de deteção de ocupação (MDO) deverá ser capaz de determinar o número de ocupantes através da energia elétrica agregada e desagregada dos equipamentos elétricos instalados numa habitação. Para determinar o número de ocupantes, o modelo será implementado recorrendo a um algoritmo de aprendizagem supervisionada e a uma base de dados contendo o consumo agregado e desagregado de energia da instalação, assim como a informação do número de ocupantes à mesma frequência do registo da energia.

1.3 Estrutura da Dissertação

Para além do primeiro capítulo onde é realizada a introdução ao trabalho realizado, enquanto a motivação e os objetivos propostos, este documento é ainda constituído pelos quatro capítulos seguintes:

- **Capítulo 2 - Estado de arte**

O segundo capítulo apresenta, em primeiro lugar, o estado de arte relativo à detecção de ocupação, referenciando as abordagens realizadas por vários autores. Em segundo lugar são apresentados estudos já realizados relativos à utilização de algoritmos para detetar ocupação através da energia elétrica. Por último, apresenta-se o conhecimento prévio relativo aos temas abordados neste trabalho.

- **Capítulo 3 - Arquitetura do Sistema**

Neste capítulo é explicada a arquitetura do sistema implementado para detetar ocupação através do consumo de energia. São também apresentados os cenários experimentais para validar a eficácia do sistema, sendo descritas as tecnologias e o modelo de consumo de energia utilizado nesta solução.

- **Capítulo 4 - Testes e Resultados**

O quarto capítulo mostra o comportamento do sistema em vários cenários, permitindo assim avaliar, em que condições, a solução desenvolvida, apresenta bons resultados na detecção do número de ocupantes numa determinada habitação.

- **Capítulo 5 - Conclusões finais e trabalhos futuros**

O capítulo final resume o trabalho realizado, comparando os resultados obtidos, em cada cenário, com os objetivos propostos. Apresentam-se também futuros trabalhos com o objetivo de melhorar as limitações do sistema desenvolvida nesta dissertação.

ESTADO DE ARTE

Este segundo capítulo descreve o estado de arte e o conhecimento prévio que constituem as linhas orientadoras desta dissertação. No subcapítulo 2.1 são descritas as abordagens existentes para detetar ocupação dando exemplos concretos. Em seguida, no subcapítulo 2.2 é feita uma breve explicação dos algoritmos de desagregação de cargas e apresentados diversos trabalhos realizados relacionados com a deteção de ocupação não intrusiva utilizando o consumo de energia. Por fim, no subcapítulo 2.3 é apresentado o conhecimento prévio que sustenta o trabalho desenvolvido nesta dissertação, designadamente conceitos importantes relacionados com técnicas de *machine learning* como a aprendizagem supervisionada, sendo descrito o algoritmo *k-nearest neighbor* e as funções utilizadas para avaliar a performance dos algoritmos de classificação.

2.1 Detecção de ocupação em edifícios

A detecção de ocupação num edifício ou habitação pode ser realizada através de uma abordagem direta ou indireta. Os métodos de detecção de ocupação diretos requerem a instalação estratégica de sensores e a participação ativa dos ocupantes, apresentando uma maior exatidão. Muitos sistemas de detecção de ocupação utilizam sensores ambientes (e.g., sensores infravermelhos passivos, sensores de portas, sensores acústicos, câmaras de vídeo, sensores de contacto, sensores de CO_2 , interruptores magnéticos, etc.). Para além disso, a utilização de sensores ambientes apresenta muitos desafios. Os sensores de CO_2 são mais sensíveis às condições exteriores, necessitando ser constantemente calibrados tendo em conta o sistema de ventilação [13]. Por outro lado, os sensores de movimento não distinguem animais de pessoas provocando falsos positivos [4, 14].

Diversos sensores, como os de movimento e de portas, podem ser usados na detecção de pessoas para fornecer informação a sistemas que utilizam termostatos inteligentes para poupar energia desligando automaticamente os sistemas de aquecimento, ventilação e ar condicionado (AVAC) quando os ocupantes estão a dormir ou fora de casa. Em [14] os autores analisaram 51 base de dados, onde em 8 destas foram instalados sensores para saber, em tempo real, quando os ocupantes estavam acordados, a dormir, chegavam ou saíam de casa. Os resultados mostraram que a instalação de termostatos inteligentes e sensores para saber o nível de ocupação, constitui uma solução bastante vantajosa para melhorar a eficiência energética das habitações, poupando em média 28% no consumo de energia.

O número e tipo de sensores existentes num sistema de monitorização de ocupação requer o equilíbrio entre precisão de detecção, custo global e complexidade do sistema. Os custos associados à manutenção de sensores, devem-se à necessidade de os calibrar regularmente e trocar as baterias [15]. Nas habitações, onde o custo com instalação de sensores para monitorização de ocupação deve ser baixo, apenas sensores baratos e de baixa precisão são utilizados, sendo a manutenção feita, na maior parte das vezes, pelos próprios ocupantes.

Por outro lado, os métodos de monitorização de ocupação indiretos são mais económicos, mas menos precisos comparando com os métodos diretos, utilizam fontes de informação contextual (e.g., localização por GPS, ligação à internet por cabo ou Wi-Fi, calendários e aplicações utilizadas pelos ocupantes, etc.). Estes métodos relacionam uma ou mais atividades com a ocupação, não são intrusivos nem é necessária a instalação e manutenção de sensores. Contudo, a precisão destes métodos depende da regularidade dos ocupantes e constante atualização das fontes de informação. Por exemplo, a utilização do sinal GPS requer que todos os ocupantes usem o smartphone dentro de casa, sendo a sua utilização limitada à duração das baterias destes dispositivos. Na prática, a localização por GPS é combinada com outras tecnologias, normalmente com a localização por wi-fi e outros sensores localizados na separação das divisões da habitação [4, 7, 11, 16–18].

Vários estudos têm sido realizados utilizando sensores na detecção de ocupação. Em

[19] foram usadas medições de sonar para determinar ocupação, os autores partiram do princípio de que o corpo humano provoca uma alteração, bem definida e distinta, nas ondas de som comparativamente ao ar e a outros objetos. Através de 20 voluntários e utilizando sistemas de áudio e vídeo instalados juntos aos computadores, os investigadores concluíram ser possível detetar com precisão os utilizadores apenas com medições de 10 segundos de duração. Dodier *et. al* [20] instalaram múltiplos sensores infravermelhos passivos em cabinets fechados e implementaram um classificador Baysiano probabilístico para inferir ocupação. Hailemariam *et. al* [21] desenvolveram um método para detetar ocupação utilizando algoritmos de classificação do tipo árvores de decisão, instalando diversos sensores de baixo custo num cubículo de trabalho para medir variáveis relacionadas com a ocupação (e.g. iluminação, som, nível de CO_2 , energia consumida pelas aplicações e movimento). Na implementação do algoritmo foram utilizadas combinações de características, introduzindo uma ou várias medições dos diversos sensores. A combinação que apresenta melhores resultados corresponde à informação recolhida apenas pelo sensor de movimento, tendo-se obtido uma precisão de 98.4%. Significa isto que introduzir informação proveniente de múltiplos sensores, a precisão global do sistema utilizando árvores de decisão não melhora, evitando-se assim a ocorrência do fenómeno de *overfitting*. Segundo os autores, outra conclusão reside no fato das árvores de decisão detetarem ocupação com melhor precisão relativamente a técnicas que utilizam um valor limiar.

Contudo existem alternativas à utilização de sensores de modo a obter informação relativa à presença humana numa habitação ou edifício. Segundo Chen *et al.* [7] o consumo energético de uma habitação muda sempre que os ocupantes estão presentes. Deste modo, será apresentada no subcapítulo 2.2 a monitorização de ocupação não intrusiva utilizando apenas medidores de energia constitui uma ferramenta poderosa.

2.2 Deteção de pessoas através do consumo de energia

O consumo de energia pode ser dividido em duas categorias distintas, de acordo com a forma como a informação é registada. Por um lado, tem-se a energia agregada que diz respeito ao consumo total da instalação. Contrariamente, é possível ter a energia desagregada, isto é, o registo em separado do consumo de todos os equipamentos da instalação.

Tendo-se apenas o consumo agregado de energia é necessário desenvolver um procedimento para desagregar a energia consumida ao nível do dispositivo, separando os diagramas de carga do diagrama de carga do consumo total da habitação.

Este processo é conhecido como monitorização não intrusiva de cargas *Non-intrusive load monitoring* (NILM) e refere-se à análise do consumo de energia e observação do estado de ativação dos dispositivos elétricos. Os sistemas que utilizam abordagens NILM tem como objetivo estudar o comportamento elétrico dos dispositivos instalados numa habitação.

Este procedimento implica a construção de um algoritmo que identifica um dispositivo através da assinatura elétrica, isto é, a característica associada ao seu comportamento elétrico. A característica elétrica refere-se a uma grandeza física aquando do funcionamento do dispositivo, como por exemplo a potência ativa ou a transformada de Fourier da forma de onda da corrente consumida. A recolha dos dados necessários à implementação do algoritmo dividem-se em dois tipos, de acordo com a frequência de amostragem. As características elétricas adquiridas a baixa frequência, isto é, abaixo de 1 Hz designam-se por características macroscópicas. Contrariamente, características adquiridas a frequências acima de 1 Hz designam-se por características microscópicas [22, 23].

Na literatura umas das primeiras abordagens NILM foi proposta por George W. Hart em 1992 enquanto investigador do *Massachusetts Institute of Technology* (MIT). O sistema desenvolvido por Hart, designado *Nonintrusive Appliance Load Monitoring* (NALM) [24], permite monitorizar os equipamentos elétricos instalados numa habitação através das variações da potência ativa e reativa. Posteriormente, Hart comparou as variações de energia consumida pelos vários equipamentos com as características dos dispositivos presentes numa base de dados, conseguindo detetar quando os dispositivos são ligados e desligados. Apesar das abordagens NILM identificarem com eficácia dispositivos de alta potencia (e.g. sistemas AVAC), a maioria delas requer conjuntos de treino detalhados tornando o processo de aprendizagem muito demorado.

A temática da deteção de ocupação não exige que se discrimine o tipo de equipamentos no consumo agregado de energia, é apenas necessário procurar evidencias, isto é, características na curva de carga que sinalizem a interação entre pessoas e equipamentos elétricos. O simples fato de ligar e desligar manualmente dispositivos como televisores, fogão, micro-ondas, chaleira ou circuitos de iluminação é sinal de que a casa está ocupada. Por outro lado, existem equipamentos como frigoríficos, congeladores ou o modo *standby* de algumas aplicações que não apresentam relação direta com ocupação.

Um dos investigadores que mais estudos tem realizado nesta área é Wilhelm Kleiminger do *Institute for Pervasive Computing* ETH em Zurique, tendo desenvolvido a base de dados *Electricity Consumption and Occupancy* (ECO) [25], onde consta o consumo de energia à granularidade de 1 Hz (i.e., 86.600 medições por dia) de 6 habitações na Suíça durante um período de 7 meses. Os autores instalaram tomadas inteligentes para medir o consumo de energia em cada dispositivo (i.e., energia ativa e reativa nas três fases) assim como sensores para detetar a presença dos utilizadores. Para registar a ocupação ao longo desse período foi utilizado um *tablet* em cada habitação. Posteriormente foi testada a performance de alguns algoritmos NILM sobre base de dados ECO. Os resultados mostraram que os métodos supervisionados apresentam melhor precisão na desagregação de energia, comprovando que a granularidade de 1 Hz é suficiente para detetar quando alguns dispositivos são ativados ou desligados.

Num dos seus primeiros estudos [16] os resultados obtidos mostraram que usando algoritmos de classificação simples é possível obter uma precisão na deteção de ocupação

acima dos 80%. Adicionalmente ao consumo de energia registado pelo medidor inteligente, foi recolhida informação através de sensores infravermelhos passivos e tomadas inteligentes. Para registar o número de ocupantes, em cada habitação, foi instalado um *tablet*.

As 10 características derivadas da curva de potência, por fase e em intervalos de 15 minutos, utilizadas na classificação foram: potência média, o desvio padrão e a soma das diferenças absolutas (i.e., o cálculo da diferença absoluta entre as medições de energia adjacente somado às seguintes medições). Os algoritmos de classificação utilizados foram o *Support Vector Machine*, *K-Nearest Neighbour*, técnicas utilizando um valor limiar e o Modelo Hidden Markov.

Num estudo posterior, Kleiminger *et al.* [17] utilizaram a base de dados desenvolvida em [16], contento energia agregada e desagregada, mas em vez de 10 características foram usadas 35. As características obtidas através da curva de potência, por fase, foram: valores absolutos, variância e dependência temporal. Em primeiro lugar, as características correspondentes aos valores absolutos abrangem ao valor mínimo, máximo e média aritmética em cada período de 15 minutos. Em segundo lugar, uma alta variância no consumo de energia pode indicar interação dos ocupantes com as cargas. Por último, os picos de energia dependem normalmente da altura do dia e, se é dia de semana ou fim-de-semana.

À semelhança do trabalho realizado em [16], os autores implementaram os mesmos algoritmos de classificação, adicionando apenas um modelo Gaussiano. Uma vez que alguns classificadores não apresentaram uma boa performance foi necessário reduzir a dimensão da base de dados. Assim sendo, os autores implementaram dois métodos de seleção de características: *Sequential Forward Selection* e *Principal Component Analysis*. A técnica *Sequential Forward Selection* é um método direto e não-paramétrico que a cada iteração escolhe a melhor característica que maximiza a performance do algoritmo, terminando o processo até todas as características serem selecionadas ou até algumas terem sido excluídas [26]. *Principal Component Analysis* corresponde a uma transformação ortogonal de um conjunto de características possivelmente relacionadas numa combinação linear de variáveis sem relação, designadas por componentes principais. Estas são ordenadas de acordo com relação à ocupação. Neste estudo, Kleiminger *et al.* obtiveram melhor performance usando o método PCA do que o SFS, devido à grande redundância entre algumas características tornando difícil escolher o melhor subconjunto. Devido à relação não linear entre as características, o classificador com melhores resultados foi o SVM.

Através da energia agregada fornecida pelo contador inteligente, foi possível alcançar, em média, uma precisão na deteção de ocupação de 92% no verão e 94% no inverno. Utilizando a energia desagregada, as classificações efetuadas apenas pelo estado de ativação dos dispositivos diretamente relacionados com a ocupação obtiveram, em uma das habitações, uma precisão de 81% no verão e 84% no inverno. De acordo com estes valores pode-se concluir que a precisão na deteção de ocupação depende muito da utilização de equipamentos com grande relação com a ocupação e aumento significativo do consumo de energia num curto período de tempo (e.g. forno, micro-ondas, secador de cabelo ou

aspirador), se este tipo de dispositivos não foram usados pelos utilizadores, naturalmente o algoritmo errará mais vezes.

Para além de métodos utilizando algoritmos de aprendizagem supervisionada, Kleiminger *et al.* desenvolveram um simples algoritmo baseado num valor limiar para detetar ocupação. O objetivo deste estudo foi avaliar a possibilidade de construir um método sem ser necessário ter a informação do número de ocupantes para treinar o algoritmo. Estes procedimentos são bastantes úteis pois é possível utilizar diretamente o consumo de energia fornecido pelas companhias de eletricidade, através das medições efetuadas pelos contadores inteligentes. Na Figura 2.1 verificam-se os períodos ocupados a vermelho e os desocupados a azul, enquanto que na parte inferior da figura apresenta-se a técnica de deteção de ocupação. No período compreendido entre as 0:00 e as 6:00 o algoritmo considera a habitação desocupada, contudo os utilizadores estão presentes, mas inativos, razão pela qual o gráfico da parte superior da Figura considere a habitação como ocupada. O algoritmo foi implementado utilizando oito variáveis (e.g. média-dia, mínimo-dia, média-noite, mínimo-noite, média-15-minutos, mínimo e máximo-15 minutos) atribuindo o estado de ocupada à habitação caso a energia durante um período de minutos ultrapasse um dado valor limiar.

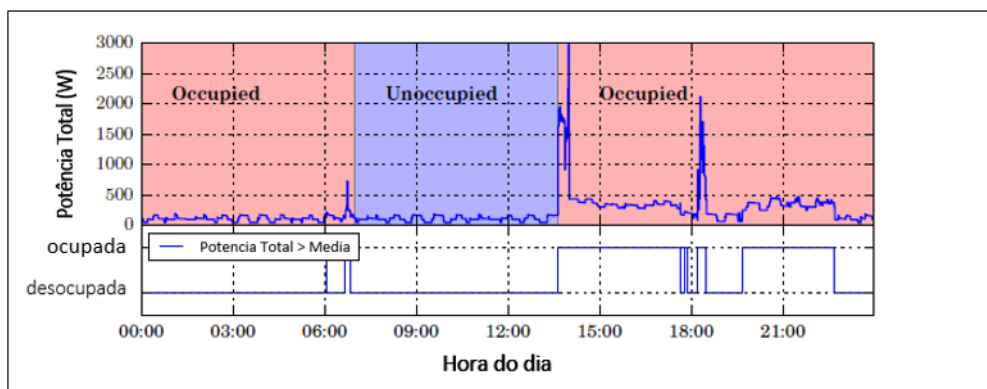


Figura 2.1: Algoritmo de deteção de ocupação baseado num método utilizando um valor limiar. Adaptado de [17].

Os resultados obtidos foram apenas aceitáveis em 1 das 5 habitações usadas no teste, sendo a melhor precisão de 78%. Em suma, apesar de ser possível detetar ocupação com um simples método threshold, a utilização de contadores inteligentes para detetar ocupação requer a utilização de métodos de aprendizagem supervisionada mais avançados, preferencialmente à frequência de 1 Hz.

Chen *et al.* em [7] exploraram o potencial da monitorização de ocupação não intrusiva (*Non-Intrusive Occupancy Monitoring* - NIOM). O objetivo desse estudo passou por inferir ocupação duas habitações a partir do consumo agregado de energia, registado por medidores inteligentes. Segundo os autores, uma habitação quando ocupada, apresenta um acréscimo em métricas como o valor do consumo médio e no desvio padrão em relação

à média, durante curtos intervalos de tempo. Desde modo, foi desenvolvido um método utilizando um valor limiar baseado no algoritmo NIOM que deteta mudanças em cada métrica. Dado que pequenas interações com cargas podem não aumentar significativamente o consumo médio de energia ou o desvio padrão, neste estudo, os autores adicionaram como característica a variação absoluta da energia. Esta variável sinaliza ocupação se a variação da energia ultrapassar, durante um determinado intervalo de tempo, o valor definido como limite. Segundo os autores, os sistemas de ar condicionado mudam significativamente o perfil de carga, pois aumentam muito rapidamente a energia consumida num curto período de tempo, constituindo um bom indicador de ocupação. Deste modo, quando o perfil de carga não apresenta picos de consumo de energia a tarefa do algoritmo utilizando um valor limiar em determinar o estado de ocupação torna-se mais complicada originando maior número de falsos negativos. A precisão na detecção de ocupação foi melhor no verão atingindo os 75.3%, pois existe uma maior tendência por parte dos utilizadores em ligar manualmente os ar condicionados e sistemas de ventilação quando chegam a casa e, a desliga-los quando saem. Este fato faz com que cargas de segundo plano (i.e. cargas que estão ligadas sem necessidade de interação humana ou que podem ser comandadas à distancia), como os ar condicionados, sejam consideradas melhores indicadores de presença no período de verão do que cargas tipicamente interativas. Estes resultados mostram que, através de uma simples abordagem, é possível detetar ocupação através da energia agregada registada pelos medidores inteligentes.

Outro estudo de ocupação binária, cujo problema, tem como solução a classificação a duas classes (e.g. ausência - presença de pessoas numa habitação), os autores de [27] utilizaram versões avançadas de algoritmos de *machine learning* como *Support Vector Machines* (SVM), Árvores de Decisão, Florestas Aleatórias e *Backpropagation*. A base de dados é composta por três conjuntos de dados: i) o consumo de energia das aplicações da instalação (e.g. equipamentos de cozinha, televisão, máquina de lavar, frigorífico e secador de cabelo), registado por tomadas inteligentes guardando a corrente de cada dispositivo a cada minuto; ii) o consumo de água, registado por quatro sensores para monitorizar todas as fontes de fornecimento de água; iii) a informação do número de ocupantes, registado através de sensores infravermelhos ativos posicionados nas portas para detetar a entrada e saída de pessoas enviando a contagem por um servidor web. Posteriormente, os dados foram agregados numa única base de dados, contendo 9 características (e.g. iluminação geral, frigorífico, televisão, máquina de lavar roupa, secador, água-quente-cozinha, água-fria-cozinha, água -máquina lavar loiça, água-máquina de lavar roupa) e a ocupação alvo correspondente (i.e. habitação com 3 ocupantes). Para seleccionar as características que apresentam melhor relação com a ocupação foi utilizada a técnica de informação mútua para medir quanto uma variável aleatória fornece informação sobre outra. Por outro lado, para otimizar os parâmetros de entrada dos classificadores foram realizadas 100 simulações de Monte Carlo e utilizadas as suas versões melhoradas dos algoritmos (i.e. através do algoritmo *AdaBoost*). Os resultados mostraram que as Florestas Aleatórias e as Árvores de Decisão foram os classificadores que melhor performance obtiveram, pela

métrica F-measure 83.37% e 82.79% respetivamente.

Em [18] Ming *et al.* utilizando base de dados, já existentes (ECO, UMass Smart e PC), contento o consumo de energia, desenvolvendo três abordagens distintas. A primeira abordagem-*Base Learning* (BL), tem como objetivo i) verificar se o consumo de energia é suficiente para inferir ocupação, e ii) verificar se a precisão na deteção de ocupação depende do tamanho da base de dados. Em primeiro lugar, os autores utilizaram um simples método de valor limiar e, em seguida, algoritmos de classificação mais avançados como SVM, Florestas Aleatórias e Árvores de Decisão. Apesar das diferenças de performance dos algoritmos dependendo da base de dados utilizada, os algoritmos de classificação mais avançados apresentaram melhores resultados, comprovando assim que a energia é um bom indicador de ocupação. Em segundo lugar, utilizando os algoritmos com melhor performance da abordagem BL, pela Figura 2.2 verifica-se que a precisão de classificação aumenta significativamente quando o conjunto de treino acresce 5% ao inicial, sendo que abaixo desse valor torna-se difícil estimar corretamente a correspondência energia-ocupação.

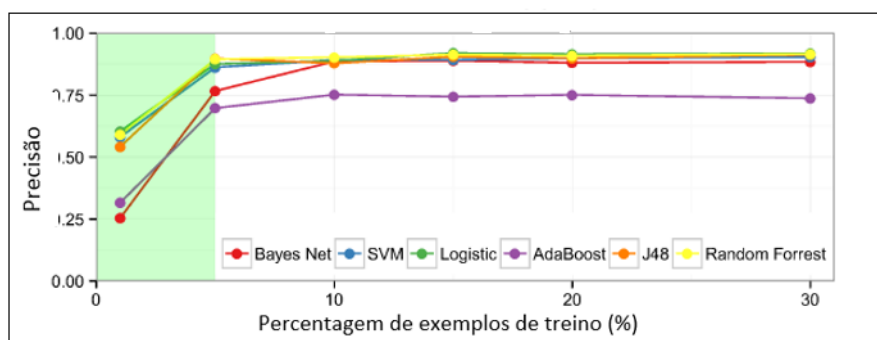


Figura 2.2: Precisão dos algoritmos variando o tamanho do conjunto de treino no intervalo (1, 5, 10, 15, 20, 30)%. Adaptado de [18].

Na segunda abordagem- *Non-intrusive learning* (NL) uma vez que não existem dados suficientes para implementar algoritmos de classificação, foram desenvolvidos dois algoritmos i) *Multiview-based Iteration* - inicializa pelo horário do calendário dos utilizadores sendo a informação de presença iterativamente melhorada até as condições de paragem serem alcançadas; ii) *Surrogate Loss* - reconhece o erro no conjunto de treino e aprende através da função de uma função de perdas. De um modo geral, os dois métodos apresentam uma performance satisfatória, apesar de ocorrerem mais erros de classificação quando a habitação está desocupada. Estes erros ocorrem essencialmente devido ao fato de os utilizadores por vezes não cumprirem o calendário à risca e programarem dispositivos para trabalhar enquanto estes não estão em casa. Estes métodos foram implementados usando os calendários dos ocupantes para determinar o estado de ocupação, tendo obtido 74% a 89% de eficácia em edifícios residenciais e cerca de 90% em comerciais.

Na terceira abordagem - *Transfer Learning* (TL) pretende-se detetar ocupação em habitações mesmo que existam poucos dados para implementar algoritmos de aprendizagem.

Deste modo, os métodos TL têm como objetivo encontrar padrões no perfil de carga da habitação alvo com outras habitações com características similares, tentando determinar a correta correspondência energia-ocupação.

Em [28] foi desenvolvido um sistema de monitorização de ocupação não intrusiva sem utilizar algoritmos de aprendizagem supervisionada, isto é, sem conhecimento dos períodos de ocupação e desocupação nem fase treino. Através das leituras do consumo de energia, medido por um contador inteligente instalado numa habitação, os autores construíram uma simples e prática abordagem de deteção de ocupação através da curva de carga. Na Figura 2.3 apresenta-se a framework SHARK desenvolvida pelos autores.

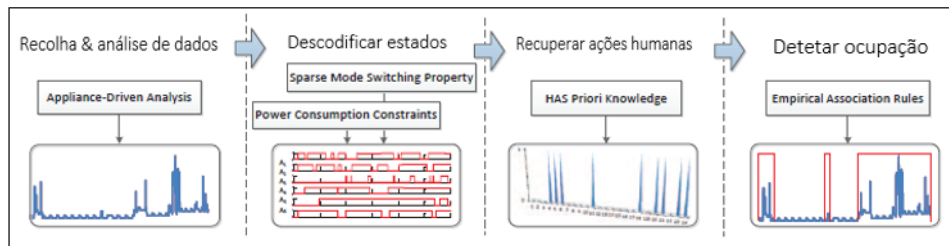


Figura 2.3: Framework SHARK para deteção de ocupação não intrusiva sem fase de treino. Adaptado de [28].

Em primeiro lugar, foi feita a recolha e análise dos dados do consumo agregado de energia, analisando as propriedades do perfil de carga das aplicações para encontrar relação com as atividades humanas. A maioria das aplicações trabalham segundo um ou múltiplos modos, apresentando cada modo um valor de energia habitual mas distinto, por exemplo uma lâmpada tem um consumo estável, contrariamente um secador pode operar em quatro modos diferentes (e.g. baixo/alto ou frio/quente). Em segundo lugar, decodificaram os modos de estado (i.e. ligado ou desligado) dos dispositivos de acordo com o seu perfil de carga a partir do consumo agregado de energia. Normalmente, os modos ligar/desligar de uma aplicação pode ser i) automático (e.g. frigorífico, congelador ou termo-acumulador); ii) manual (e.g. secador, TV); iv) semi-automático (e.g. o micro-ondas pode ser ligado pelos utilizadores mas pode ser desligado automaticamente), assim sendo os autores construíram uma base de dados contendo todos estes eventos. Em terceiro lugar, foram recuperadas as ações humanas usando conhecimento à priori a partir do registo de atividade da relação entre humano e os eventos ligar/desligar aplicações. Por último, foram inferidos os estados de ocupação da habitação a partir das ações humanas recuperadas, numa primeira fase os autores dividiram o consumo de energia em intervalos de 15 minutos, sendo a deteção de ocupação da habitação realizada pelas regras: i) se o número de ações é maior que o valor limiar; ii) se ocorrerem eventos ligar/desligar durante os períodos de ocupação/desocupação, definidos pelos utilizadores, a casa pode ser considerada ocupada ou desocupada mesmo que os eventos indiquem o contrário de modo prevenir falsos positivos e falsos negativos. Os resultados obtidos mostraram que mesmo sem usar abordagens de aprendizagem supervisionada, isto é, sem existir uma

fase de treino e informação de ocupação à priori, a framework desenvolvida pode detetar ocupação com alta precisão.

Em [29] os autores demonstraram que o consumo de energia apresenta um potencial elevado na deteção de ocupação. À semelhança de outros estados arte, foram utilizados algoritmos de classificação como KNN e SVM, assim como HMM para analisar a dependência temporal da ocupação. SVM é um algoritmo muito usado em reconhecimento de padrões devido, não só ao fato de gerar fronteiras de decisão não-lineares, assim como permitir uma grande margem ao classificador.

2.3 Conhecimento prévio

2.3.1 Conceitos Fundamentais

Para inferir ocupação a partir da curva de carga é necessário mapear uma função, a partir do conjunto de características (e.g. consumo médio de 100 watts nos últimos 15 minutos), para as classes de ocupação como presente-ausente ou o número de ocupantes habituais (e.g. [característica] \rightarrow classe). Esta função de mapeamento, designada por classificador é determinada no processo de classificação utilizando algoritmos de aprendizagem supervisionada. Assim, a partir dos dados de entrada o classificador aprende, usando posteriormente essa aprendizagem para classificar novos dados.

2.3.2 Aprendizagem Supervisionada

Na aprendizagem supervisionada infere-se o classificador através de um conjunto de dados mapeado para duas ou mais classes, i.e., conhece-se o valor de entrada e da correspondente saída. O objetivo pretendido com as técnicas de *machine learning* é encontrar a correta relação entre entrada e saída, normalmente designada por modelo.

Os dados devem ser divididos em conjuntos de treino e teste para evitar o fenómeno de *overfitting*, i.e., construir um classificador que descreve o ruído nos dados em vez da relação entre características e classes. O conjunto de teste é utilizado para verificar a capacidade do classificador em mapear corretamente os novos dados de entrada às saídas correspondentes.

2.3.3 K-nearest neighbor

O K-nearest neighbor é um algoritmo simples, robusto, versátil e fácil de entender. É amplamente utilizado para determinar a saída de uma dada função, quer em problemas de classificação ou regressão. Os classificadores do tipo KNN usam modelos não paramétricos na classificação, i.e., não existe uma fase de aprendizagem propriamente dita. Em vez disso, durante a classificação de uma instância no conjunto de teste, o algoritmo KNN em primeiro lugar encontra os k exemplos mais próximos no conjunto de treino de acordo com uma medida de distância (e.g., Euclidiana, Hamming, Manhattan ou Minkowski).

Através dos exemplos de classes conhecidas desses k vizinhos, o algoritmo escolhe a classe pela maioria dos votos das classes desconhecidas. Nesta dissertação foi utilizado o classificador *KNeighborsClassifier* da biblioteca *scikit-learn* [30].

2.3.3.1 Pseudocódigo do KNN

A Figura 2.4 mostra o pseudocódigo do algoritmo KNN. Em primeiro lugar é importado o conjunto de treino- X e as classe- Y correspondentes e é escolhido o valor k . Em cada iteração é calculada a distância em toda a amostra- x , determinando o conjunto I com as k distâncias mais próximas. Por fim, será escolhida a classe com mais representantes dentro do conjunto I .

```

Início:
Classificar ( $X, Y, x$ ) //  $X$ : conjunto de treino,  $Y$ : classes de  $X$ ,  $x$ : amostra desconhecida
for  $i = 1$  to  $m$  do
    Determinar a distância  $d(X_i, x)$ 
end for
Determinar a distância do conjunto  $I$  contendo os índices das  $k$  menores distâncias
 $d(X_i, x)$ 
return classe maioritária para  $\{Y_i \text{ onde } i \in I\}$ 
Fim.

```

Figura 2.4: Pseudocódigo do algoritmo KNN

2.3.4 Avaliação

A classificação da ocupação de uma habitação é normalmente um problema a duas classes. Durante um intervalo de tempo de 15 minutos, uma habitação pode estar ocupada ou desocupada. Para avaliar os modelos de classificação, podem ser usadas métricas [31] como a exatidão, precisão, *recall* e *F-measure*, retirados a partir da matriz de confusão representada na Tabela 2.1. A matriz de confusão permite verificar se o classificador previu bem uma determinada classe, analisando o número de verdadeiros positivos e negativos, assim como o número de falsos positivos e falsos negativos. A diagonal da matriz de confusão permite verificar o número de pontos corretamente classificados para cada classe, assim se o modelo de classificação tiver 100% de exatidão a matriz terá uma diagonal pura, isto é, todos os pontos classificados com a classe correta.

Tabela 2.1: Matriz de Confusão

		Classe Actual (Valor Real)		Total $VP + FP$ $FN + VN$ N
		p (ocupada)	n (desocupada)	
Classe Prevista	p' (ocupada)	Verdadeiro Positivo	Falso Positivo	
	n' (desocupada)	Falso Negativo	Verdadeiro Negativo	
Total		$VP + FN$	$FP + VN$	

A correta classificação de uma habitação como **ocupada** é designada por **verdadeiro positivo**. Contrariamente, a correta classificação de uma habitação como **desocupada** é designada por **verdadeiro negativo**. A incorreta classificação de uma habitação, como **ocupada** e **desocupada** é, respetivamente, designada por **falso positivo** e **falso negativo**. Através destes conceitos, podem ser usadas diversas métricas para avaliar a performance dos classificadores que serão apresentadas nas secções seguintes.

2.3.4.1 Falsos Negativos e Falsos Positivos

Atribuir a uma habitação o estado de **desocupada**, quando de fato esta está ocupada, constitui uma classificação como **falso negativo**. A taxa de falsos negativos de um classificador, descrita pela Equação 2.1, é dada pelos falsos negativos a dividir por todos os intervalos de tempo desocupados (verdadeiros e falsos negativos).

$$TFN = \frac{fn}{fn + vn} \quad (2.1)$$

Da mesma forma, classificar uma habitação como **desocupada**, quando esta está na realidade ocupada, diz-se que ocorreu um **falso positivo**. A taxa de falsos positivos de um classificador, descrita pela Equação 2.2, é dada pelos falsos positivos a dividir por todos os intervalos de tempo ocupados (verdadeiros e falsos positivos).

$$TFP = \frac{FP}{FP + VP} \quad (2.2)$$

2.3.4.2 Exatidão

A exatidão refere-se à proximidade das medidas ao valor correto, sendo dada pelo quociente entre verdadeiros positivos e verdadeiros negativos e o total de elementos da amostra (i.e. verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos) e descrita pela Equação 2.3. Na classificação com múltiplas classes, o resultado desta função diz respeito à exatidão com que o algoritmo atribui aos novos pontos a classe correta.

$$Exatidao = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.3)$$

2.3.4.3 Precisão

A precisão, descrita pela Equação 2.4, é dada pelo quociente entre os verdadeiros positivos e a soma dos verdadeiros positivos e falsos positivos. A precisão é assim vista como a capacidade do classificador em não atribuir o valor verdadeiro a um novo ponto que é negativo, mostrando a percentagem dos resultados que é relevante, ou seja, os positivos previstos que são realmente positivos verdadeiros.

$$Precisao = \frac{VP}{VP + FP} \quad (2.4)$$

2.3.4.4 Recall

Por outro lado, o *recall*, descrito Equação 2.5, é dada pelo quociente entre os verdadeiros positivos e soma dos verdadeiros positivos e falsos negativos. Esta métrica refere-se à habilidade do classificador em encontrar todos os exemplos positivos, determinando a percentagem total dos resultados corretamente classificados pelo algoritmo, ou seja, os positivos verdadeiros que são bem classificados.

$$Recall = \frac{VP}{VP + FN} \quad (2.5)$$

Contudo não é possível descrever a eficiência de um classificador utilizando apenas precisão e o *recall*. Combinando estas duas métricas, a função *F-measure*, dada pela Equação 2.6, é definida pela média harmónica entre precisão e *recall*.

$$F - measure = \frac{2 * (Precisao * Recall)}{Precisao + Recall} \quad (2.6)$$

Quando o resultado desta métrica é perto de 1, implica melhor combinação entre precisão e *recall*, contrariamente valores mais baixos significa pior exatidão ou precisão ou ambos.

ARQUITETURA DO SISTEMA

Neste capítulo é explicada a arquitetura do sistema para detetar ocupação através do consumo de energia recorrendo a um algoritmo de aprendizagem supervisionada e também o modo como esta arquitetura foi implementada. Em primeiro lugar, no subcapítulo 3.1 é descrito o modelo conceptual do modelo de deteção de ocupação (MDO), apresentando o seu modelo funcional e arquitetural. Em seguida, no subcapítulo 3.2 é explicada a implementação de todo o sistema, como as tecnologias utilizadas, os sistemas de gestão de dados e deteção de ocupação. Neste sub-capítulo apresenta-se ainda a camada que permite avaliar e visualizar a ocupação inferida a partir do MDO.

3.1 Modelo Conceptual

O modelo conceptual apresenta o conceito que sustenta o desenvolvimento do MDO, isto é, a descrição de alto nível referente à organização e funcionamento do sistema, desde a fonte de informação e tratamento dos dados, até ao algoritmo de detecção de ocupação. Este modelo é constituído pelo modelo funcional, o modelo arquitetural e o modelo de dados. O modelo funcional apresenta as funcionalidades do MDO. O modelo arquitetural mostra a relação entre os vários blocos do sistema. Para finalizar, o sistema de gestão de dados serve de base à construção do MDO.

3.1.1 Modelo Arquitetural

A arquitetura do MDO, representada na Figura 3.1, é composta por três camadas principais, sendo cada uma destas dividida em vários módulos e camadas secundárias. A primeira camada diz respeito à gestão dos dados, estabelecendo a comunicação com o MCE necessário à implementação de todos os sistemas. A segunda camada refere-se à detecção de ocupação, recebendo os dados tratados e prontos para serem utilizados pelos módulos que constituem o SDO. Por fim, na terceira camada são novamente efetuadas operações sobre os resultados obtidos pela camada anterior, sendo possível avaliar e visualizar os resultados obtidos.

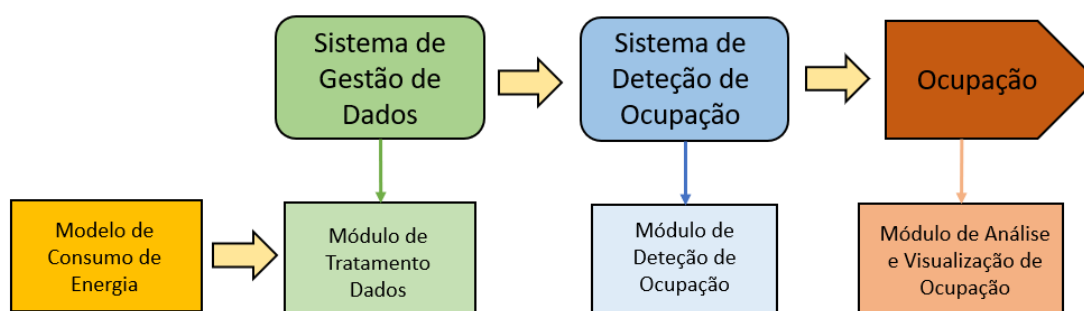


Figura 3.1: Modelo Arquitetural do Sistema

3.1.1.1 Sistema de Gestão de Dados

O Sistema de Gestão de Dados, representado na Figura 3.2, é a camada responsável pelo tratamento, organização e armazenamento dos dados. Esta camada utiliza o MCE como base de dados onde se obtém os diagramas de cargas (i.e. contendo a energia agregada e desagregada) e a informação do número de ocupantes.

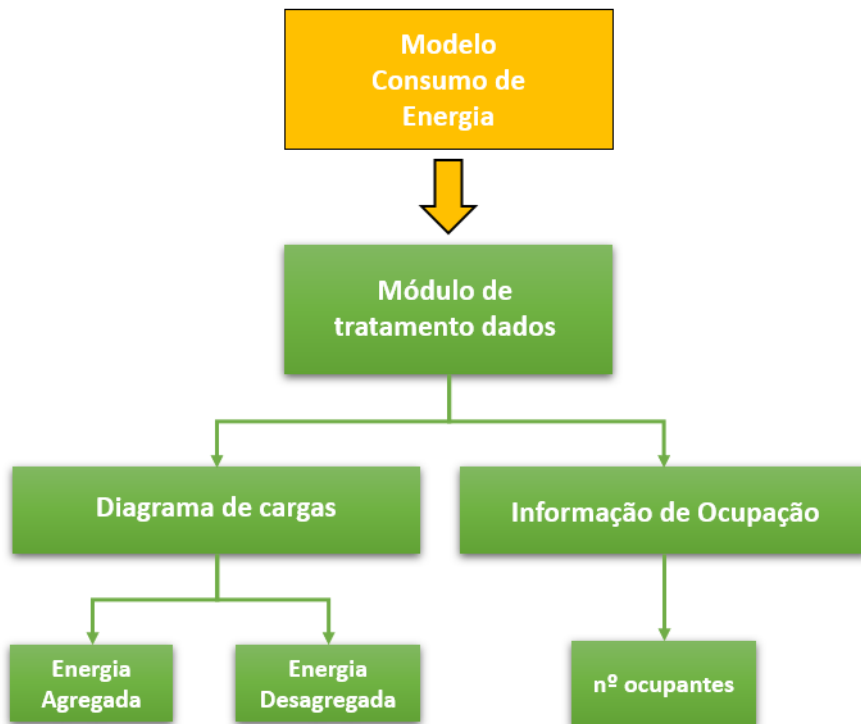


Figura 3.2: Arquitetura do Sistema de Gestão de Dados

Esta camada serve de base ao processo de aprendizagem implementado no SDO, apresentado na secção seguinte. É então necessário armazenar o consumo de energia agregada e desagregada dos dispositivos elétricos e o número de ocupantes. O consumo agregado de energia corresponde ao somatório do consumo individual de cada equipamento (i.e. consumo desagregado de energia).

3.1.1.2 Sistema de Detecção de Ocupação

Esta camada é responsável por detetar o número de ocupantes utilizando um algoritmo para classificar novos dados, sejam eles a energia agregada ou desagregada dos dispositivos elétricos. O sistema de detecção de ocupação, representado na Figura 3.3, é composto pelo módulo de detecção de ocupação, dividido em três módulos: módulo de aprendizagem, módulo de classificação e módulo de avaliação. O módulo de detecção de ocupação interage com a camada de dados para obter toda a informação necessária de modo a realizar o processo de aprendizagem e construir um modelo que seja capaz de classificar corretamente a habitação como ocupada e desocupada, assim como identificar o número de ocupantes num determinado intervalo de tempo.

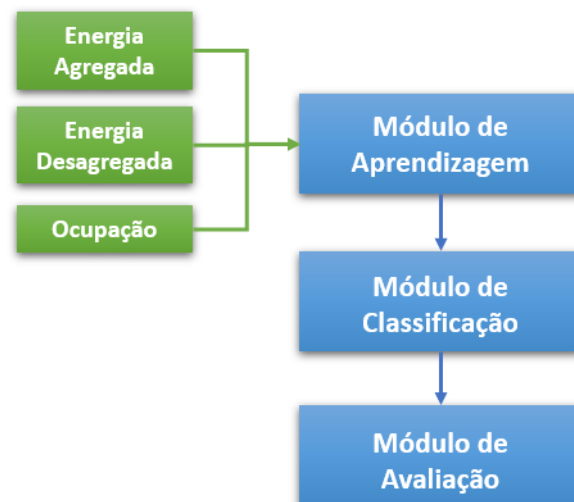


Figura 3.3: Arquitetura do Sistema de Gestão de Dados

3.1.1.3 Ocupação

Esta camada tem como objetivo tratar os dados provenientes da camada SGD de modo a avaliar e visualizar os resultados obtidos, comparando graficamente o número de ocupantes real com o número de ocupantes determinados pelo MDO.



Figura 3.4: Arquitetura da camada Ocupação

Na Figura 3.4 verifica-se que a camada Ocupação é constituída pelo módulo de análise e visualização de ocupação (MAV). O módulo da média móvel é responsável por aplicar um filtro.

3.2 Implementação

3.2.1 Tecnologias Utilizadas

A implementação do MDO foi realizada utilizando a aplicação cliente-servidor Jupyter Notebook [32]. Através desta aplicação é possível desenvolver código utilizando linguagens de programação como R ou Python e ainda, criar elementos de texto como equações, figuras ou tabelas. O desenvolvimento das várias camadas da arquitetura do MDO foi realizado através da linguagem de programação Python. Utilizou-se Python pois é uma linguagem de alto nível, de script e orientada a objetos, permitindo implementar de forma simples todas as camadas desta arquitetura. Para desenvolver código em Python utilizando a aplicação Jupyter Notebook foi necessário instalar a distribuição open-source Anaconda [33] num computador portátil.

Foi decidido utilizar ficheiros do tipo Excel para guardar todos os dados relativos às simulações efetuadas. Para analisar, preparar e gerar dados foram usadas as bibliotecas Numpy e Pandas que permitem construir facilmente estruturas de dados e realizar estatísticas sobre dados em larga escala, utilizando a linguagem Python. Na implementação do algoritmo de classificação recorreu-se à ferramenta Scikit-learn que utiliza também várias bibliotecas do Python como Pandas, Numpy e Matplotlib para visualização de dados.

3.2.2 Modelo de Consumo de Energia

Uma vez que não foi possível recolher dados no laboratório, utilizando um contador inteligente, recorreu-se ao modelo doméstico de consumo de eletricidade desenvolvido por Richardson *et al.* em [34] como base de dados. Este modelo permite gerar sinteticamente a partir do ficheiro Excel disponibilizado pelos autores, com uma resolução de 1 minuto (i.e. ou à frequência de 1 Hz), o consumo de energia e o número de ocupantes em uma habitação. Os dados simulados pelo modelo combinam padrões de ocupação ativa (i.e. quando os utilizadores estão em casa e acordados) e perfis diários de atividade (i.e. que descrevem a maneira como os utilizadores ocupam o seu tempo realizando determinadas atividades).

A arquitetura deste modelo, representada na Figura 3.5, é composta por vários blocos. Os perfis diários de atividade são comuns a todas as habitações e representam a probabilidade dos utilizadores realizarem uma determinada tarefa ao longo de 24 horas. A cada habitação está associada um conjunto de dados de ocupação ativa e um conjunto de aplicações instaladas. Cada uma das aplicações tem a si associada um perfil de atividade diário, assim quando um ou mais dispositivos elétricos são ligados cada uma das suas características da energia consumida é usada para calcular o consumo de energia total da habitação.

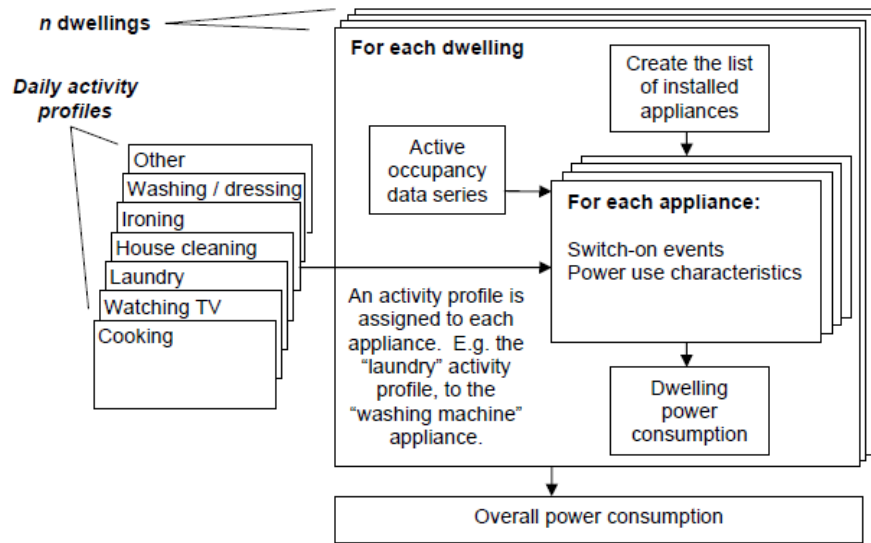


Figura 3.5: Arquitetura do modelo de consumo de energia [34].

O perfil de atividade diária está diretamente relacionado com o número de ocupantes ativos (i.e. o modelo permite simular entre 1 e 5 ocupantes), dia-de-semana ou fim-de-semana e, ainda o mês do ano. O modelo permite também alocar, de forma automática ou manual, equipamentos elétricos à simulação do consumo de energia. A Tabela 3.1 apresenta os 15 equipamentos escolhidos para desenvolver um método de deteção de ocupação, sendo composta pelas colunas: CMT ou seja, o tempo médio em minutos do ciclo de trabalho dos equipamentos elétricos; PMCT referente à potência média do ciclo de trabalho em Watts e a coluna Categoria que distingue os equipamentos de acordo com a sua relação com a ocupação. A escolha destes equipamentos deve-se ao fato de serem os mais usualmente utilizados nas habitações em Portugal e ainda, pela grande maioria apresentar forte relação com a ocupação. A categoria A diz respeito a equipamentos como televisões, chuveiro elétrico e aplicações ligadas à cozinha, que apenas trabalham na presença dos ocupantes e são prontamente desligados quando estes não se encontram em casa ou estão a dormir. Por outro lado, a categoria B representa os equipamentos que são ligados na presença dos ocupantes mas que podem continuar a operar quando estes saem de casa ou estão a dormir, como por exemplo as máquinas de lavar roupa ou da loiça. Por último, os equipamentos da categoria C podem trabalhar independentemente da presença ou ausência dos ocupantes, como é o caso dos frigoríficos ou difusores de sinal Wi-Fi.

Tabela 3.1: Equipamentos elétricos e sua relação com a ocupação.

Equipamento	CMT (min)	PMCT (Watts)	Stand-by (Watts)	Categoria
Frigorífico	22	190	0	C
Wi-Fi	60	100	9	C
Ferro de Engomar	30	1000	0	A
Aspirador	20	2000	0	A
Computador	300	141	5	A
Televisão 1	73	124	3	A
Televisão 2	73	124	3	A
Fogão	16	2400	1	A
Forno	27	2125	3	A
Microondas	30	1250	2	A
Chaleira Elétrica	3	2000	1	A
Máquina da Loiça	60	1131	0	B
Máquina de Lavar Roupas	138	406	1	B
Chuveiro Elétrico	3	9000	0	A
Iluminação	-	-	-	A

3.2.3 Sistema de Gestão de Dados

O modelo apresentado na secção anterior permite apenas simular consumo de energia e ocupação para uma habitação durante 24 horas. Contudo, para construir um modelo de classificação, com o objetivo de detetar de ocupação, utilizando um algoritmo de aprendizagem supervisionada, é necessário um conjunto de dados para treino e teste extenso. Deste modo, foi necessário efetuar o equivalente a 100 dias de consumo de energia e ocupação.

3.2.3.1 Tratamento dos dados

O código do procedimento responsável pelo tratamento e simulação do consumo de energia e o número de ocupantes encontra-se representado na Figura 3.6. Em primeiro lugar, abre-se o ficheiro Excel do modelo doméstico de consumo de eletricidade e, através de um ciclo, em cada iteração, são executadas as macros da ocupação (i.e. responsável por simular o número de ocupantes durante 24 horas) e das aplicações (i.e. responsável por simular o consumo de energia dos equipamentos previamente alocados). Em segundo

lugar, copia-se os dados referentes ao consumo de energia dos equipamentos e o número de ocupantes da habitação simulados para uma nova folha de Excel.

O ciclo é terminado após se obter o equivalente a 100 dias de consumo de energia e correspondente informação de ocupação simulados. A organização dos dados na folha do Excel passou por adicionar, em cada iteração, 1440 linhas correspondentes a 24 horas de consumo de energia e número de ocupantes à frequência de 1 minuto. Desde modo, cada folha Excel terá 144000 linhas (i.e. 100 dias), constituindo uma base de dados suficientemente grande para treinar o modelo de aprendizagem. Posteriormente, é apenas simulado 1 dia para testar a performance do modelo em detetar o número de ocupantes quando surgem novos dados de consumo de energia.

```
### Abrir o Ficheiro Excel
wb = xw.Book(r'C:\Users\Diogo Oliveira\Anaconda3\Tese\Notebooks\model_teste_10 equip.xlsx')

### Começar o ciclo
i = 12
f = 1451
s = 2
count=1
while count < 101:

    ### Correr a macro da ocupação
    ExcelMacro = wb.macro('Sheet1.btnRunOccupancy_Click')
    ExcelMacro()
    time.sleep(2)

    ### Correr a macro das aplicações
    ExcelMacro = wb.macro('Sheet1.btnRunApplianceModel_Click')
    ExcelMacro()

    # Fridge freezer
    wb.sheets['data_10_1_jan'].range("A1").value = 'Fridge freezer'
    source_rng1 = "I" + str(i) + ":I" + str(f)
    dest_rng1 = "A" + str(s)
    my_values = wb.sheets['appliance_sim_data'].range(source_rng1).options(ndim=2).value
    wb.sheets['data_10_1_jan'].range(dest_rng1).value = my_values
```

Figura 3.6: Procedimento para gerar os dados de consumo de energia e ocupação

3.2.4 Sistema de Deteção de Ocupação

O Sistema de Deteção de Ocupação (SDO) é composto pelo módulo de aprendizagem, módulo de classificação e pelo módulo responsável pela avaliação do modelo.

3.2.4.1 Módulo de Aprendizagem

O módulo de aprendizagem, representado na Figura 3.7, é responsável pelo carregamento dos dados e implementação do modelo de classificação.


```

# DataFrame com 16 características, ft contem as características selecionadas
ft=[0,1,2,3,4,5,6,7,8,9,10,11,12,13,14]
features = pd.read_excel('model_teste_3_pessoas.xlsx', sheet_name='data_15_jan', usecols=ft)

# Data Frame com a classe alvo, tg contem as classes selecionadas
tg = [16]
target = pd.read_excel('model_teste_3_pessoas.xlsx', sheet_name='data_15_jan', usecols=tg)

# Cria um instancia vazia
scaler = StandardScaler()

# Calcula a média o desvio padrão para mais tarde normalizar as características
scaler.fit(features)

# Efetua a normalização, centrando e redimensionando usando o método transform
scaled_features = scaler.transform(features)

# Guarda o resultado como data frame
features_feat = pd.DataFrame(scaled_features)

# Divide a base de dados em conjunto de treino e teste, 70% p/ treino e p/ 30% teste
X_train, X_test, y_train, y_test = train_test_split(features_feat, target, test_size=0.3)

# Cria um classificador KNN ()
knn = KNeighborsClassifier(n_neighbors=5)

# Treina o modelo usando o conjunto de treino
knn.fit(X_train, y_train)

# Preve a resposta no conjunto de teste
y_pred = knn.predict(X_test)

```

Figura 3.7: Módulo de aprendizagem do modelo K-Nearest Neighbor

Em primeiro lugar, as características (i.e. da energia agregada e desagregada consumida pelas cargas elétricas da habitação) e as classes correspondentes (i.e. o número de ocupantes que interagem com as cargas elétricas) são importadas do ficheiro Excel e guardadas numa estrutura de dados, isto é, para uma *data frame*) disponibilizada pela biblioteca *Pandas* [35]. Como tal são utilizadas estruturas diferentes para guardar cargas elétricas e a ocupação. Em segundo lugar, recorrendo à biblioteca *Scikit-learn*, procede-se à normalização dos dados, através da função *scaler.transform()* para que estejam todas na mesma ordem de grandeza. Em terceiro lugar, a base de dados é dividida, através da função *train_test_split()* em 70% para conjunto de treino e os restantes 30% para o conjunto de teste. Este método tem como parâmetros as características e classes correspondentes, assim como a percentagem definida para o conjunto de teste, tendo-se optado por 30% uma vez ser o valor mais comum na literatura. Por último, é criado um classificador utilizando o algoritmo k-Nearest Neighbor através da função *KNeighborsClassifier()* [30], definindo como único parâmetro o número de vizinhos *K*. Para escolher *K* testou-se a exatidão do modelo para valores ímpares entre 1 e 9, pois este parâmetro funciona como um critério de desempate na escolha da classe do novo ponto a classificar. Na Figura 3.8 é possível verificar a exatidão de classificação dos vários modelos para os meses de Janeiro e Junho, usando na fase de treino dados relativos ao mês de Janeiro, tendo três ocupantes

e 15 equipamentos elétricos.

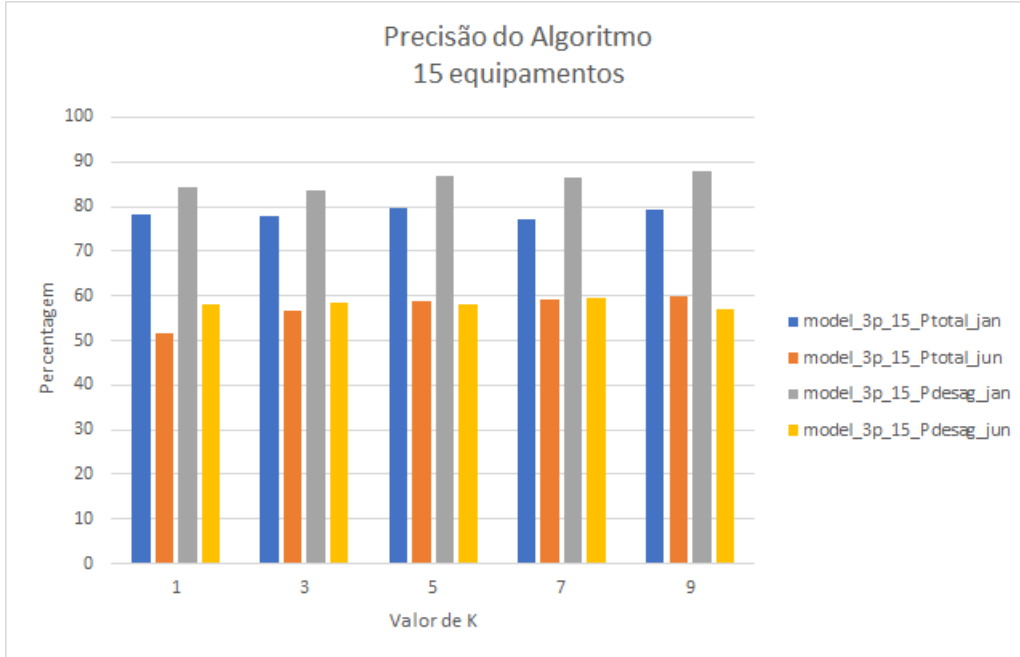


Figura 3.8: Exatidão dos modelos variando o valor de K

Analisando a figura anterior verifica-se que o modelo deteta o número de ocupantes com uma exatidão a ronda os 80% para todos os valores de k num determinado dia de Janeiro. Contrariamente, a exatidão do modelo situa-se entre os 50% e os 60% para todos os valores de k para o mês de Junho, não sendo neste caso muito fiável. Assim sendo, o k escolhido será o valor 5 para a implementação dos cenários experimentais apresentados na secção 4.1 do Capítulo 4.

A medida utilizada na implementação dos classificadores *knn* para determinar os K vizinhos mais próximos foi a distância que por defeito a biblioteca *scikit-learn* propõe, neste caso a distância Euclidiana dada pela Equação 3.1.

$$d(x, x') = \sqrt{(x_1 - x_1')^2 + (x_2 - x_2')^2 + \dots + (x_n - x_n')^2} \quad (3.1)$$

3.2.4.2 Módulo de Classificação

O módulo de classificação, representado na Figura 3.9, segue os primeiros dois passos do módulo de aprendizagem. São importados os dados relativos à energia consumida pelos equipamentos elétricos, necessários como parâmetro de entrada do classificador *knn*, utilizando o método *predict()* para determinar o número de ocupantes, guardando o resultado no vector *y_pred_day*. É também guardado numa *data frame* o número de ocupantes real para avaliar, no módulo de avaliação, a performance do módulo de classificação.

```

# DataFrame contendo 15 características, ft2 contem as características selecionadas - JANEIRO
ft2=[2,3,4,5,6,7,8,9,10,11,12,13,14,15,16]
features2 = pd.read_excel('model_teste_3_pessoas.xlsx', sheet_name='day_15_jan', usecols=ft2)

# Data Frame com a classe alvo, tg2 contem as classes selecionadas
tg2 = [18]
target2 = pd.read_excel('model_teste_3_pessoas.xlsx', sheet_name='day_15_jan', usecols=tg2)

# Cria um instancia vazia
scaler2 = StandardScaler()

# Calcula a média o desvio padrão para mais tarde normalizar as características
scaler2.fit(features2)

# Efetua a normalização, centrando e redimensionando usando o método transform
scaled_features2 = scaler.transform(features2)

# Guarda o resultado como data frame
features_feat2 = pd.DataFrame(scaled_features2)

# Preve a resposta do modelo em novos dados de consumo de energia (durante 24 horas)
y_pred_day = knn.predict(features_feat2)

```

Figura 3.9: Módulo de classificação do modelo K-Nearest Neighbor

3.2.4.3 Módulo de Avaliação

O módulo de avaliação avalia performance da classificação quando surgem novos dados de consumo de energia. Para realizar esta avaliação recorreu-se à função `metrics.accuracy_score()` sendo passado como parâmetros o número de ocupantes real e a ocupação obtido pelo modelo de classificação. Para além disso, utiliza-se a função `classification_report()` para medir a qualidade da classificação do algoritmo. Esta função permite assim analisar o número de classificações verdadeiras e falsas (i.e. determinar os VP, VN, VN e FN). A análise mais detalhada do resultado da classificação, pode ser obtida através da matriz de confusão. Utilizando a função `confusion_matrix()` é possível verificar o resultado da classificação para cada classe.

```

from sklearn.metrics import confusion_matrix

print("Exatidão para o conjunto de teste:", metrics.accuracy_score(target2, y_pred_day))
print(classification_report(target2, y_pred_day))
print(confusion_matrix(target2, y_pred_day))

```

Figura 3.10: Módulo de avaliação do modelo K-Nearest Neighbor

3.2.5 Visualização de Ocupação

Nesta camada foi necessário atenuar a constante variação, em curtos períodos de tempo, do número de ocupantes determinados pelo SDO e representados na Figura 3.11. Para tal, recorreu-se à média móvel como técnica de alisamento da série temporal, funcionando como um filtro passa-baixo, eliminando as oscilações visíveis no gráfico da ocupação detetada.

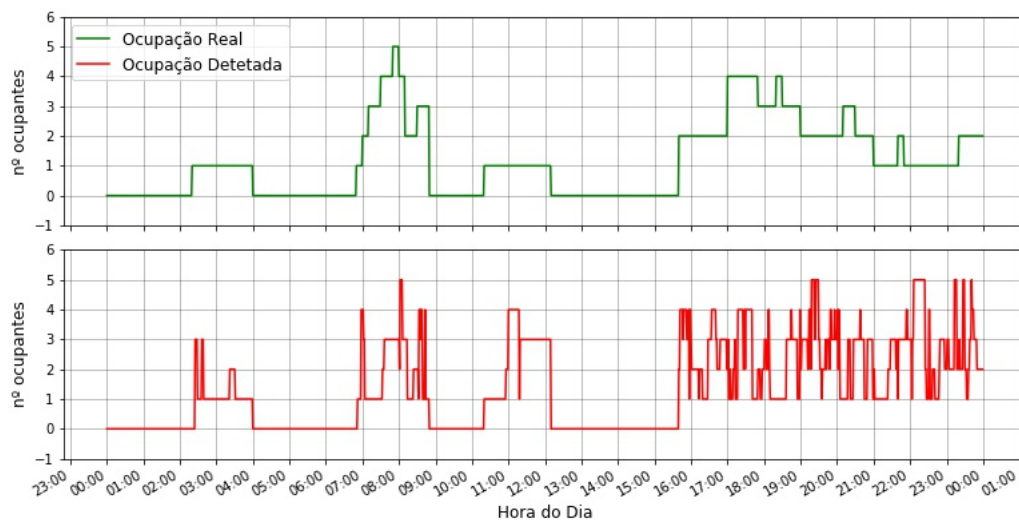


Figura 3.11: Resultado da aplicação do algoritmo antes de aplicar a média móvel

Esta técnica passa por construir uma nova série a partir da média de N elementos consecutivos da série original. Iniciando-se pelos N primeiros valores, introduz-se assim um atraso (i.e. o valor do intervalo N expressa-se em função dos valores presentes e passados). O processo repete-se avançando sucessivamente até ao fim da série, atenuando os picos da nova série do cálculo da média dos N elementos da série.

Sabendo que, quanto maior for o intervalo N , mais suave e menos aleatório será o gráfico da ocupação, em contrapartida maior será o desvio em relação ao gráfico da ocupação real. Deste modo, a escolha do intervalo N foi realizada testando a performance do SDO nos períodos de 5, 10, 15 e 20 minutos, sendo que generalidade dos casos, uma janela de 5 minutos é a que mais se aproxima do valor de ocupação real. A média móvel do número de ocupantes foi implementada usando o código da Figura 3.12.

```
aux = ocup.drop('time_jan', 1)
rolling_mean = aux.rolling(window=5).mean()
rolling_mean_round = rolling_mean.round()
rolling_mean_round_final = rolling_mean_round.fillna(method='backfill', limit=4)
```

Figura 3.12: Módulo da média móvel do nº de ocupantes

Em primeiro lugar, aplica-se a função `rolling_mean().mean()` definindo o tamanho do intervalo no qual se pretende aplicar a média móvel. Em segundo lugar, arredonda-se o número de ocupantes para valores inteiros. Por fim, preenche-se os primeiros índices, usados para calcular a primeira média móvel, através da função `fillna()`. Como se pretende preencher os N valores anteriores, no método-*method* deve ser utilizado o parâmetro *backfill*, assim caso $N = 5$ então o limite-*limit* deve ter o valor 4.

Pela Figura 3.13 constata-se que após aplicar a média móvel a curva do número de ocupantes apresenta menos picos, sendo mais próxima da curva original.

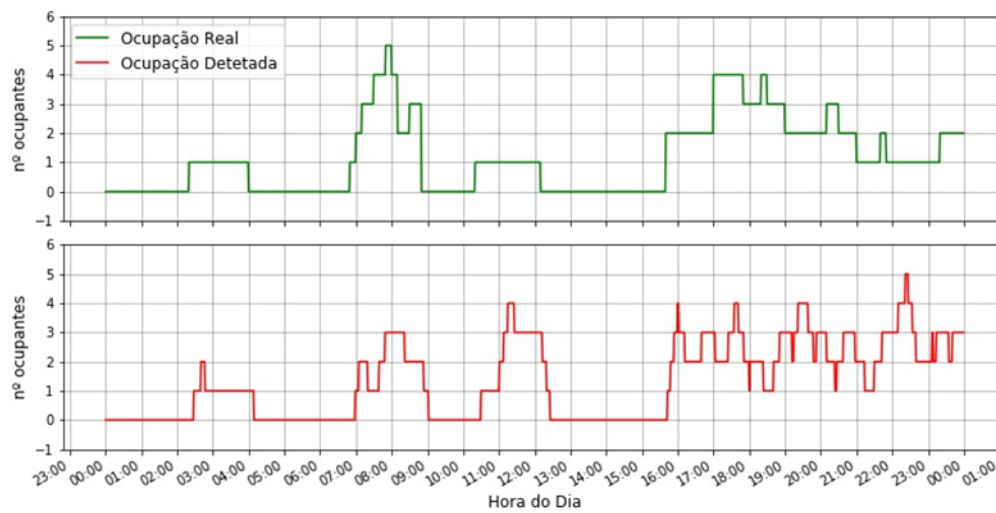


Figura 3.13: Resultado da aplicação do algoritmo após aplicar a média móvel

TESTES E RESULTADOS

No presente capítulo são apresentados os cenários experimentais, procedendo-se de seguida à avaliação do funcionamento do MDO através da análise dos testes experimentais realizados em todos os cenários. Deste modo, a avaliação do MDO é efetuada analisando cada uma das suas camadas através de métricas como a exatidão, precisão e *recall*.

Primeiramente procedeu-se à validação do Módulo de Aprendizagem (MA), uma vez que inicialmente o sistema não tem conhecimento que características da curva de carga da habitação melhor indicam ocupação. Como tal, o MDO aprendeu através da energia agregada e desagregada padrões nos dados introduzidos no MA que permitem construir um modelo capaz de detetar corretamente o número de ocupantes na habitação.

Seguidamente ao introduzir novos dados de consumo de energia e, utilizando o classificador do MA implementado, foi construído um Módulo de Classificação (MC) capaz de detetar o número de ocupantes tendo apenas disponível a energia agregada ou desagregada consumida numa determinada habitação. Deste modo, tendo disponível um conjunto de dados foi realizada a média aritmética da precisão alcançada pelo MC durante 10 dias contendo o consumo de energia de uma determinada habitação. Por último, de modo a avaliar com maior clareza o comportamento do MDO, foi escolhido a partir desse conjunto de dados o dia em que o MC melhor detetou o número correto de ocupantes.

Posteriormente, o resultado da classificação obtido no MC foi enviado para o Módulo de Análise e Visualização (MAV), avaliando o MDO depois de aplicada a média móvel (sub-capítulo 3.2.5) ao resultado da classificação. O atraso utilizado foi de 5 minutos de modo a filtrar os picos do gráfico da ocupação detetada. Nesta fase, a avaliação do MDO foi realizada pormenorizadamente, analisando o relatório de classificação e a matriz de confusão.

4.1 Cenários Experimentais

De modo a verificar experimentalmente a capacidade do MDO em detetar o número de ocupantes serão considerados quatro cenários. Através dos cenários 1, 2 e 3 são avaliadas as variáveis de entrada utilizadas na fase de construção do algoritmo (i.e. mês do ano, nº de ocupantes, nº de equipamentos, energia agregada e desagregada). O cenário 4 utilizará a melhor combinação de variáveis testadas nos cenários 1, 2 e 3 para avaliar o tipo de equipamentos que devem ser utilizados pelo MDO.

Em todos os cenários são apenas utilizados dados do consumo de energia relativos a dias-de-semana, sendo que nos cenários 1 e 2 consideram-se sempre 3 ocupantes e varia-se o número de equipamentos. Nos sub-cenários 1.2 e 2.1 utilizam-se todos os 15 equipamentos descritos na Tabela 3.1, enquanto que nos sub-cenários 1.2 e 2.2 reduz-se para 10 equipamentos usados (i.e. Frigorífico, Wi-Fi, Ferro de Engomar, Computador, Televisão 1, Fogão, Micro-ondas, Máquina de Lavar Roupa, Chuveiro Elétrico e Iluminação), por último nos sub-cenários são usados apenas 5 equipamentos (i.e. Frigorífico, Computador, Televisão 1, Fogão e Iluminação).

Pretende-se com estes dois cenários verificar se é mais vantajoso utilizar a energia agregada ou desagregada. No cenário 1 os dados utilizados no módulo de aprendizagem são referentes apenas ao mês de Janeiro, pretendendo-se testar o comportamento do MDO para os meses de Janeiro e Junho. Por outro lado, no cenário 2 conjugam-se dados relativos aos meses de Janeiro e Junho na entrada do módulo de aprendizagem, sendo testado o comportamento do MDO apenas para o mês de Junho.

Contrariamente aos cenários 1 e 2, pretende-se com o cenário 3 verificar a capacidade do MDO em detetar corretamente o número de ocupantes quando existem mais ocupantes ativos na habitação. Assim sendo, no cenário 3 utiliza-se apenas a energia desagregada referente ao mês de Janeiro à entrada do módulo de aprendizagem, sendo sempre usados 10 equipamentos elétricos, variando o número de ocupantes (i.e. 1, 3 e 5).

Por último, o cenário 4 constitui uma otimização dos cenários 1, 2 e 3, avaliando-se assim a influência da relação dos vários tipos de equipamentos elétricos com a ocupação. Em primeiro lugar, serão descartados a energia despendida pelo frigorífico e modem de Wi-Fi, uma vez que apresentam um ciclo de trabalho constante, não contribuindo significativamente para o aumento do consumo de energia quando os utilizadores estão em casa. Em segundo lugar, analisando os dados do consumo desagregado de energia, verificar-se-á que equipamentos apresentam um maior consumo de energia, utilizando esses dados como características do MDO.

4.2 Cenário 1

Este primeiro cenário mostra que através da energia agregada (i.e. somatório da energia consumida individualmente por cada um dos 15 equipamentos da habitação) foi obtida uma exatidão de 75.5% no MA e, em média, uma exatidão de 66.38% e 55.45% no MC,

quando se pretende detetar o número de ocupantes tendo como informação diagramas de carga pertencentes a Janeiro e Junho, respetivamente. A primeira coluna da Tabela 4.1 identifica os sub-cenários, a segunda o número de equipamentos usados como características de entrada do algoritmo, a terceira coluna indica o mês do ano dos novos dados a serem classificados e as três últimas colunas a exatidão do MA, MC e MAV. Comparando os resultados anteriores, com a exatidão obtida no MA, utilizando a energia desagregada em cada um dos sub-cenários, conclui-se ser mais vantajoso utilizar a energia desagregada como característica de entrada dos módulos de aprendizagem e classificação. A partir dos equipamentos escolhidos neste cenário como características do MA, verifica-se que a exatidão do MA é tanto melhor quanto mais equipamentos forem utilizados na fase de treino do algoritmo. Analisando separadamente os sub-cenários 1.1.1 e 1.1.2, 1.2.1 e 1.2.2, 1.3.1 e 1.3.2 é notória melhor exatidão de classificação quando o mês dos dados do consumo de energia (i.e. em que se pretende detetar ocupação) introduzidos à entrada do MC coincide com o mês dos dados utilizados na fase de treino do MA. Em média, nos sub-cenários estudados o MC apresenta uma exatidão de aproximadamente 70% em novos dados de Janeiro comparando com os 67% de novos dados pertencentes a Junho. Por último, analisando os resultados obtidos no MAV, verifica-se que o resultado de classificação é similar em todos os sub-cenários, sendo o modelo mais eficiente quando o mês do MA corresponde ao mês do MC.

Tabela 4.1: Exatidão do MA, MC e MAV utilizando a energia desagregada - Cenário 1

Cenário	nº equip.	Mês	MA	MC	MAV
1.1.1	15	Janeiro	89.69%	69.83%	69.08%
1.1.2		Junho		66.17%	65.65%
1.2.1	10	Janeiro	87.84%	70.06%	69.19%
1.2.2		Junho		66.10%	65.65%
1.3.1	5	Janeiro	86.89%	71.06%	70.11%
1.3.2		Junho		67.51%	66.85%

A Figura 4.1 mostra a exatidão de classificação do MAV em cada sub-cenário. O diagrama de caixa ou *BoxPlot* é dividido em vários quartis: quartil inferior **Q1** (i.e. contém 25% dos menores valores), o valor mediano ou segundo quartil **Q2** e o quartil superior **Q3** (i.e. 75% dos menores valores). A caixa representa 50% de todos os valores registados. O *whisker* estende-se na vertical para indicar a variabilidade fora do quartil inferior e superior, ligando o limite inferior (LI) e o limite superior (LS). Estes limites são determinados até serem encontrados o valor mínimo/máximo ou até atingir uma vez e meia a distância entre **Q3** e **Q1**. A partir destes limites, todos os casos são considerados valores discrepantes ou *outliers*. Relativamente ao valor mediano, os cenários 1.2.1 e 1.3.1 apresentam o valor mais elevado, rondando os 70% em metade dos dias analisados. O cenário 1.2.1 classifica com maior eficácia dois terços dos casos pois o limite do quartil superior

apresenta uma exatidão próxima dos 75%. Apesar de nos dias selecionados este modelo atingir um máximo de 82%, num dos dias apresentou uma exatidão de classificação bastante baixa 43%. Também o cenário 1.3.1 apresenta dois casos distintos, possuindo dois *outliers* com 85% e 47% de exatidão. A caixa nos cenários 1.1.2, 1.2.2 e 1.3.2 mostra uma maior variação da exatidão de classificação em 50% dos dias estudados, apresentando valor discrepantes com resultados de classificação bastante baixos. O cenário 1.1.1 demonstra ser o modelo com melhores garantias de sucesso na classificação, pois apesar de conter um *outlier* com 47% de exatidão, apresenta o melhor resultado entre os sub-cenários, apresentando na maioria dos casos uma exatidão de 70%, obtendo na melhor das hipóteses 84%. Pela caixa, a exatidão em 50% dos dias analisados varia entre os 66% e os 75%, sendo o LI muito próximo do quartil inferior.

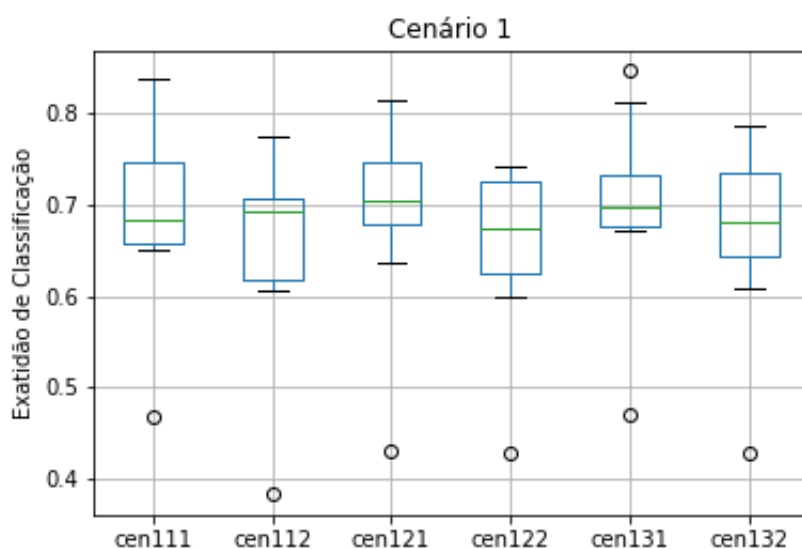


Figura 4.1: Diagrama de caixa referente à exatidão do MDO

Através do modelo implementado no sub-cenário 1.1.1, foi determinada a ocupação da habitação. De modo a estudar o comportamento do modelo, na classificação das várias classes, recorreu-se ao relatório de classificação [31] da Tabela 4.2. Pelo suporte é possível verificar que a base de dados está desequilibrada, pois em 1440 instâncias, 980 correspondem à classe 0, 420 à classe 1, 40 à classe 2 e nenhuma instância referente à classe 3.

Tabela 4.2: Relatório de Classificação

	Precisão	Recall	F-measure	Suporte
0	98%	98%	98%	980
1	90%	54%	68%	420
2	9%	40%	15%	40
3	0%	0%	0%	0

Tendo em conta todas as classes, a exatidão total obtida foi aproximadamente 84%, enquanto que através da métrica *F-measure* se obteve 87% na média ponderada entre precisão e *recall*. Pela matriz de confusão da Tabela 4.3, os resultados obtidos na classe 0 mostram precisão e *recall* de 98%, significando a boa performance do MDO pelo baixo número de FP e FN. A classificação da classe 1 também apresenta uma precisão elevada devido ao baixo número de FP. Contudo o maior número de FN levou a um baixo *recall*, isto é, o MDO classifica muitos dos novos pontos, pertencentes à classe 1, com outras classes, nomeadamente a classe 2.

Tabela 4.3: Matriz de Confusão

		Classe Prevista			
		0	1	2	3
Classe Actual	0	962	13	5	0
	1	16	228	151	25
	2	0	13	16	11
	3	0	0	0	0

De acordo com as métricas de avaliação, o MDO apresenta em geral uma elevada exatidão de classificação, contudo pela Figura 4.2 existem diferenças significativas entre os gráficos da ocupação real e detetada. Estes erros podem ser identificados pelas oscilações durante os períodos: entre 06h45 e 11h00, entre 18h00 e 19h00 e, por fim, entre 20h00 e 23h00.

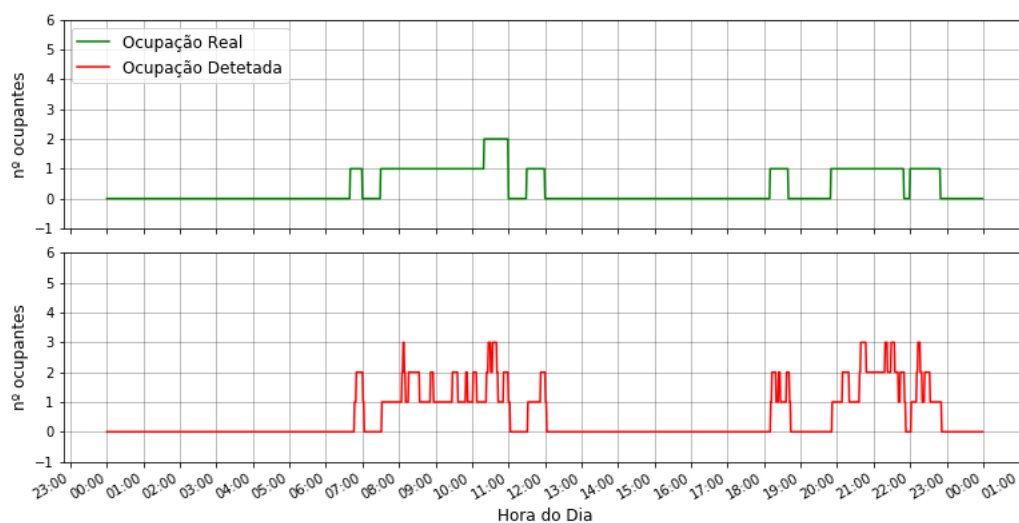


Figura 4.2: Ocupação Real e Ocupação Detetada - Cenário 1.1.1

Pela curva de carga da Figura 4.3 é possível identificar o ciclo de trabalho do frigorífico, através dos degraus com valor máximo de 227 watts. Esta informação permite concluir que o consumo de energia do frigorífico sozinho não implica presença humana. Entre as 05h45 e 07h30, a curva de carga mostra um ligeiro aumento do consumo de energia relacionado ao modo standby de alguns dispositivos (e.g. televisão, computador e fogão). Contudo, o consumo de energia neste período deve-se principalmente ao ciclo de trabalho do frigorífico e circuitos de iluminação. Tendo em conta os equipamentos usados no processo de aprendizagem, o MDO classificou, durante alguns minutos, a habitação ocupada com 2 pessoas quando efetivamente apenas 1 pessoa estava presente. Durante sensivelmente 3 horas, o MDO errou com grande frequência a classificação de novos pontos pertencentes à classe 1 e 2. Entre as 08h00 e 10h20 o consumo de energia deveu-se principalmente à utilização do computador, iluminação e frigorífico, deste modo o modelo detetou muitas vezes 2 e 3 ocupantes em vez de apenas 1. A partir das 10h20 o consumo de energia aumenta ligeiramente devido à iluminação, surgindo como indicação de que estariam presentes 2 ocupante. Apesar disso, o MDO não teve capacidade de detetar corretamente o número de ocupantes, variando com frequência o resultado da classificação (i.e. entre 1, 2 e 3 ocupantes). No final do dia, depois das 18h00 o MDO deteta com eficácia os períodos de desocupação, contudo quando existe variações no consumo de energia, o modelo erra com frequência o número de ocupantes na habitação. Com o dia analisado, é possível concluir que quando existe variação do consumo de energia devido à iluminação, computador ou televisão, o MDO classifica, na maioria das vezes, o número mínimo de ocupantes, errando sempre por excesso.

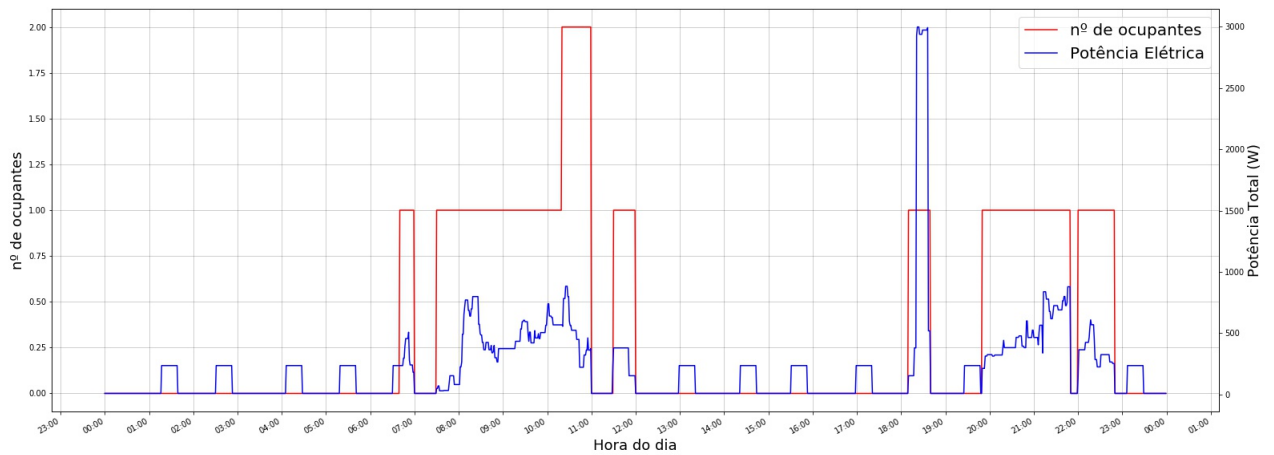


Figura 4.3: Curva de carga e ocupação real

4.3 Cenário 2

Neste cenário pretende-se avaliar a exatidão dos módulos MA, MC e MAV quando as características e as classes (i.e o diagrama de cargas e a informação de ocupação) usadas para treinar o algoritmo de classificação contém informação pertencente a vários meses. Este teste permite assim avaliar se é possível determinar com exatidão o número de ocupantes quando a base de dados de consumo de energia e ocupação não contém informação suficiente para um dado mês. Pela Tabela 4.4, através da energia agregada o MA obteve uma exatidão de 75.10%, enquanto que em média, o MC atingiu 60.27% de exatidão quando se pretende detetar o número de ocupantes tendo novos diagramas de cargas referentes ao mês de Junho. À semelhança do cenário 1 a energia desagregada apresenta melhores resultados na determinação do número de ocupantes utilizando 15, 10 e 5 equipamentos na fase de treino do algoritmo de classificação. Por outro lado, em média também o MC classifica com maior exatidão utilizando a energia desagregada comparativamente à energia agregada. É possível concluir que a exatidão do MA é tanto melhor quanto mais equipamentos forem utilizados como características no treino do algoritmo de classificação. Contudo relativamente à avaliação do MC, não existe uma diferença no resultado da classificação de novos dados quando se utiliza um maior número de equipamentos de modo a detetar ocupação. Por fim, comparando os resultados obtidos neste cenário com o cenário 1, verifica-se pela exatidão obtida na classificação do número de ocupantes que não existe vantagem em construir um classificador a partir de dados referentes a vários meses do ano. Assim sendo, pela análise dos cenários 1 e 2 conclui-se que para obter bons resultados de classificação é necessário utilizar módulos de aprendizagem onde o mês utilizado na fase de treino do algoritmo coincida com o mês dos novos a serem analisados.

Tabela 4.4: Exatidão do MA, MC e MAV utilizando a energia desagregada - Cenário 2

Cenário	nº equip.	Mês	MA	MC	MAV
2.1	15	Junho	87.63%	64.16%	63.67%
2.2	10		84.34%	64.65%	64.22%
2.3	5		83.50%	64.94%	64.61%

O diagrama de caixa da Figura 4.4 permite verificar a exatidão do MAV no cenário 2. Quanto ao valor mediano, nos sub-cenários 2.1 e 2.3 a exatidão de classificação centra-se nos 65% enquanto que o sub-cenário 2.2 apresenta um valor mais baixo. Pelo comprimento caixa, o sub-cenário 2.3 apresenta uma menor variação de classificação (i.e. entre 61% e 70%) em metade dos dias estudados, possuindo uma exatidão máxima de 77% e um valor discrepante mínimo de apenas 45%.

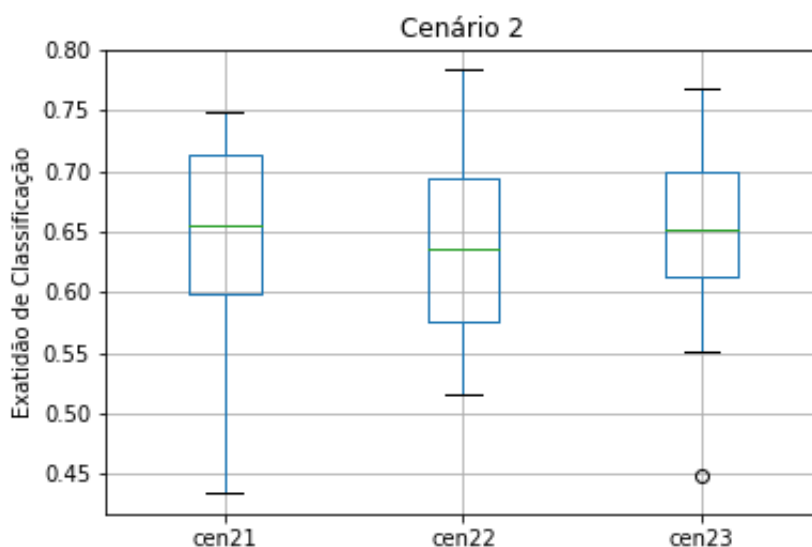


Figura 4.4: Diagrama de caixa referente à exatidão do MDO

Posteriormente, utilizando o modelo implementado no sub-cenário 2.3 foi calculado o número de ocupantes a partir do consumo de energia referente a um dia do mês de Junho. O relatório de classificação da Tabela 4.5 permite analisar os resultados obtidos pelo classificador em cada classe. De acordo com o suporte a base de dados está desequilibrada pois em 1440 instâncias, 800 correspondem à classe 0, 250 à classe 1, 310 à classe 2 e 80 à classe 3.

Tabela 4.5: Relatório de Classificação

	Precisão	Recall	F-measure	Suporte
0	88%	99%	93%	800
1	55%	74%	63%	250
2	69%	39%	50%	310
3	29%	7%	12%	80

A exatidão obtida no conjunto de todas classes foi de 77%, por outro lado pela média ponderada entre precisão e *recall*, i.e. a função *F1-measure* resultou em 74%. De acordo com a matriz de confusão da Tabela 4.6, verifica-se uma boa performance do MDO na classificação da classe 0, tendo menos casos de FN comparativamente ao número de FP. Deste modo, o *recall* de 99% indica que MDO classificou apenas 6 novos pontos da classe 0 como classe 1. Por outro lado, as restantes classes apresentam um erro relativamente maior, nomeadamente devido ao grande número de FP da classe 1 e FN da classe 2.

Tabela 4.6: Matriz de Confusão

		Classe Prevista			
		0	1	2	3
Classe Actual	0	794	6	0	0
	1	51	186	7	6
	2	60	121	120	9
	3	0	27	47	6

No geral, o MDO deteta com eficácia o número de ocupantes em determinados períodos do dia, contudo como mostram as oscilações do gráfico da ocupação detetada da Figura 4.5 existem diferenças significativas, nomeadamente entre as 06h30 e 08h15 e, entre as 15h00 e 23h00.

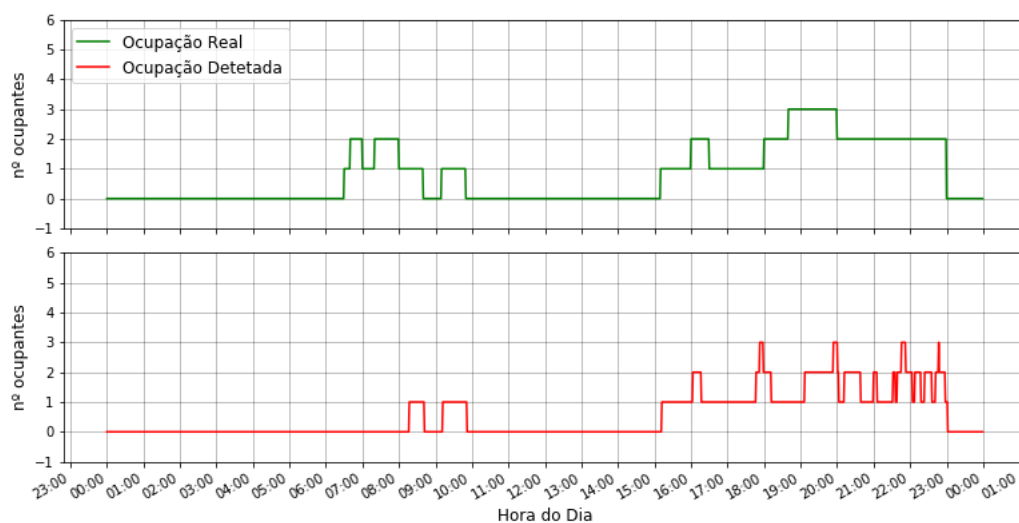


Figura 4.5: Ocupação Real e Ocupação Detetada - Cenário 2.3

De acordo com o gráfico da ocupação real, entre as 06h30 e 08h15, na habitação estiveram ativos entre 1 a 2 ocupantes, contudo o MDO considerou esta sempre desocupada. Pela curva de carga da Figura 4.6 é possível observar que durante este período o consumo de energia deveu-se principalmente ao ciclo de trabalho do frigorífico e ao modo *stand-by* do computador, televisão e fogão. Neste caso, os utilizadores não interagiram ativamente com as cargas, assim o MDO não encontrou informação no consumo de energia que indicasse ocupação, apesar de o gráfico da ocupação real mostrar que efetivamente estavam 2 ocupantes ativos na habitação. A curva de carga mostra a utilização do computador, primeiramente durante 25 minutos (i.e. entre as 08h15 e 08h39), sendo corretamente identificado pelo MDO 1 ocupante. Analisando diretamente na base de dados o consumo de energia e a curva de carga, verificou-se que a partir do momento em que o computador é desligado, o consumo de energia deveu-se apenas aos modos *stand-by* de alguns equipamentos. Também durante este período o MDO classificou corretamente o número de ocupantes habitação. Quando novamente o computador é ativado, subindo o consumo de energia, o MDO volta a detetar corretamente 1 ocupante. Durante estes 40 minutos de funcionamento do computador, precisamente às 09h30, surge um pico de consumo de energia de 100 watts devido à iluminação. Esta informação constitui um importante indício de ocupação ativa, mostrando a importância deste tipo de equipamentos na deteção de ocupação. Sensivelmente entre as 11h e 15h, o consumo de energia resultou apenas do ciclo de trabalho do frigorífico e ao modo *stand-by* dos restantes equipamentos, assim sendo a habitação corretamente considerada desocupada. A partir das 15h10, com o aumento do consumo de energia devido à utilização do computador, a habitação é novamente considerada ocupada. Contudo, o MDO apresenta dificuldades na correta identificação do número de ocupantes devido à constante variação do consumo de energia e ao maior número de equipamentos utilizados.

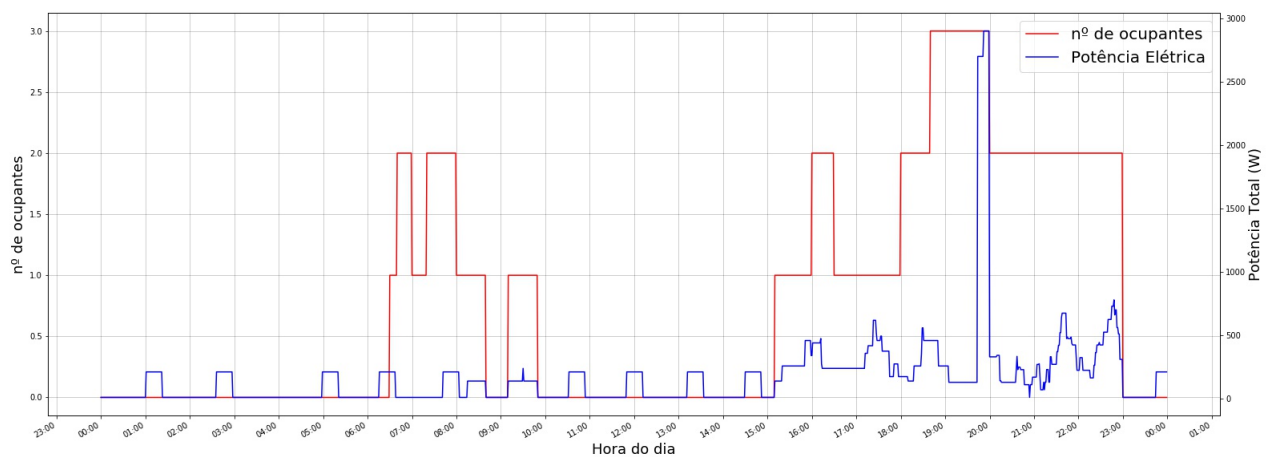


Figura 4.6: Curva de carga e ocupação real

4.4 Cenário 3

Este cenário pretende avaliar a exatidão dos módulos MA, MC e MAV mantendo fixo o número de características (i.e. 10 diagramas de carga) mas variando o número de classes (i.e. 5, 3 ou 1 ocupantes). Deste modo, foram testados 6 sub-cenários e, utilizando apenas a energia desagregada, foram implementados três MA, sendo posteriormente testados utilizando novos diagramas de cargas pertencentes aos meses de Janeiro e Junho. Estes testes permitem estudar a capacidade em detetar corretamente ocupação utilizando os classificadores KNN, usados para construir os vários módulos do MDO quando o número de classes é elevado. De acordo com os resultados obtidos, apresentados na Tabela 4.7, duas conclusões podem ser feitas: (i) a exatidão do MA é tanto maior quanto menos ocupantes a habitação possuir, como exemplo disso, o MA apresenta 98% de exatidão quando a habitação tem apenas 1 ocupante, constituindo assim um caso de classificação binária; (ii) contrariamente aos cenários anteriores, onde existia uma diferença significativa no resultado da classificação quando o mês dos novos dados difere do mês dos dados utilizados na fase de treino do algoritmo, neste cenário as diferenças são mínimas.

Tabela 4.7: Exatidão do MA, MC e MAV utilizando a energia desagregada - Cenário 3

Cenário	nº equip.	nº pessoas	MA	Mês	MC	MAV
3.1.1	10	5	81.10%	Janeiro	58.73%	57.88%
3.1.2				Junho	61.35%	61.13%
3.2.1	10	3	87.49%	Janeiro	73.40%	72.59%
3.2.2				Junho	68.82%	67.98%
3.3.1	10	1	98.34%	Janeiro	98.18%	97.41%
3.3.2				Junho	99.32%	98.66%

A Figura 4.7 mostra o diagrama de caixa com a exatidão obtida em cada sub-cenário, sendo notória a influencia do número de ocupantes no resultado final da classificação. Quando a habitação tem no máximo 5 ocupantes, a classificação em metade dos dias estudados varia entre 55% e 65% no sub-cenário 3.1.1 e, entre 53% e 51% no sub-cenário 3.1.2. No que ao valor mediado diz respeito, a exatidão situa-se, respetivamente, em 57% e 58%, nos sub-cenários 3.1.1 e 3.1.2, enquanto o valor máximo da classificação foi de 73% e 75%. Analisando os sub-cenários 3.2.1 e 3.2.2, confirma-se que os valores mínimos, máximos e medianos aumentam. Por outro lado, a variação da classificação em metade dos dias diminui, mostrando assim uma maior performance do MDO em diferentes perfis de carga. Os sub-cenários 3.3.1 e 3.3.2 pretendiam mostrar o comportamento do MDO quando uma habitação tem apenas 1 ocupante, constituindo um problema de classificação binária. Nestes sub-cenários, o intervalo da exatidão de classificação, representado pelo comprimento da caixa, é bastante pequeno, assim sendo o classificador apresenta uma boa performance na maioria dos dias analisados. Deste modo, é possível concluir que o tipo de algoritmo implementado é sensível ao número de classes, apresentando melhores resultados quanto menos classes necessitar de atribuir.

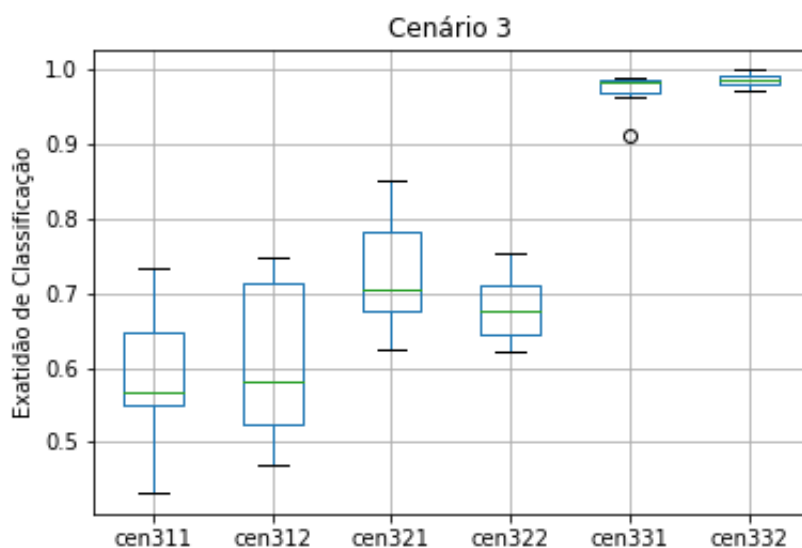


Figura 4.7: Diagrama de caixa referente à exatidão do MDO

Recorrendo ao modelo implementado no sub-cenário 3.1.1, determinou-se a ocupação de uma habitação com 5 pessoas, durante um determinado dia de Janeiro. Para analisar o comportamento do modelo na classificação de todas as classes foi necessário recorrer ao relatório de classificação da Tabela 4.8. À semelhança dos cenários anteriores, a base de dados está desequilibrada, pois existe uma maior probabilidade de a habitação ser classificada com a classe 0 do que com o somatório da probabilidade das restantes classes.

Tabela 4.8: Relatório de Classificação

	Precisão	Recall	F-measure	Suporte
0	95%	99%	97%	930
1	61%	28%	39%	240
2	0%	0%	0%	0
3	43%	38%	40%	150
4	11%	13%	12%	110
5	0%	0%	0%	10

De um modo geral, a exatidão foi de 74%, o mesmo valor foi obtido através da função *F1-measure*. A matriz de confusão da Tabela 4.9 mostra a boa performance do modelo na classificação de novos pontos pertencentes à classe 0 e, contrariamente, um elevado erro na classificação das outras classes. Enquanto o número de FP na classe 0 foi apenas 45, a classificação da classe 1 resultou em 44. Este erro pode ser identificado pelas oscilações do gráfico da ocupação detetada no período compreendido entre as 16h30 e 17h20, onde o número de FN foi igual a 27. Também a classe 3 apresentou um elevado número de FN entre as 17h20 e 19h50, ou seja, durante 1 hora e 30 minutos a deteção de ocupação da habitação não foi realizada de forma eficiente.

Tabela 4.9: Matriz de Confusão

		Classe Prevista					
		0	1	2	3	4	5
Classe Actual	0	919	8	2	1	0	0
	1	45	68	49	31	42	5
	2	0	0	0	0	0	0
	3	0	3	15	57	66	9
	4	0	31	20	45	14	0
	5	0	2	8	0	0	0

Apesar das métricas indicarem um resultado aceitável, as diferenças entre os gráficos da ocupação real e detetada, representados na Figura 4.8, mostram que o modelo determina com maior facilidade a classe 0 comparativamente às outras classes. Os erros de classificação mais significativos ocorreram no período da manhã, entre as 06h00-10h00 e entre as 16h30 e 23h00, sendo representados pelas oscilações no gráfico da ocupação detetada.

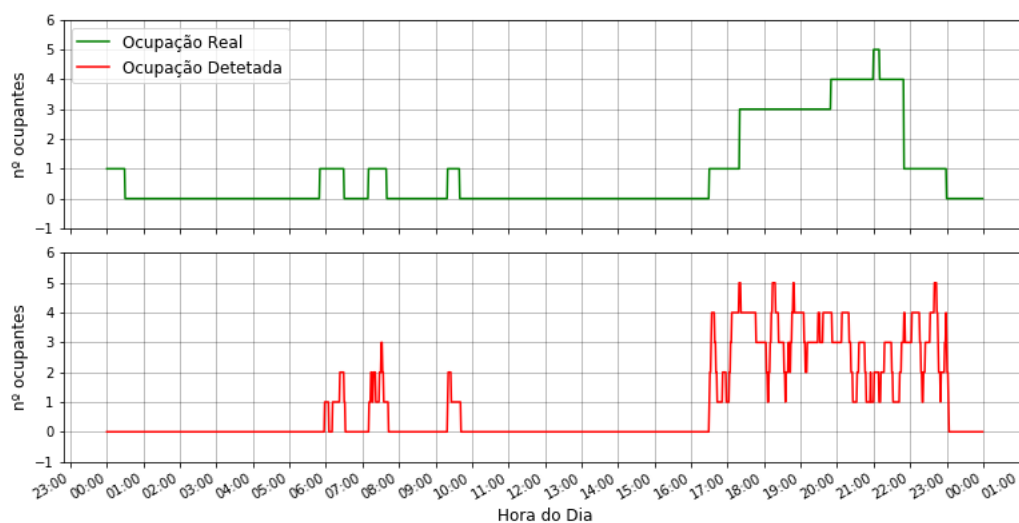


Figura 4.8: Ocupação Real e Ocupação Detetada - Cenário 3.1.1

Durante o período compreendido entre as 05h45 e 07h30, a curva de carga da Figura 4.9 apresenta um pequeno aumento no consumo de energia devido, não só ao modo *stand-by* de outros equipamentos (e.g. televisão, fogão e micro-ondas), mas principalmente relacionado com os circuitos de iluminação, computador e ciclo de trabalho do frigorífico. A partir dos equipamentos usados como características do MA, o MDO classificou, em alguns minutos, a habitação ocupada com 2 e 3 pessoas em vez de apenas uma. Entre as 07h30 e 09h20, quando o consumo de energia foi baixo (i.e. frigorífico e modo *stand-by* de alguns equipamentos), o MDO classificou corretamente a habitação como desocupada. Entre as 09h20 e 10h30, com a utilização do computador e televisão, nem sempre o MDO classificou a habitação com o número correto de ocupantes. Depois das 10h30, o baixo consumo de energia, proveniente do frigorífico e dos restantes equipamentos em modo *stand-by*, levou o MDO a classificar corretamente a habitação como desocupada. Entre as 16h30 e 23h00, a contínua utilização do computador, televisão, iluminação, fogão e micro-ondas levou a uma aumento significativo do consumo de energia, indicando a presença de vários utilizadores ativos na habitação. Consequentemente, à semelhança do início da manhã, o MDO apresentou sérias dificuldades em reconhecer o número correto de ocupantes, oscilando com grande frequência entre 1 e 5 pessoas.

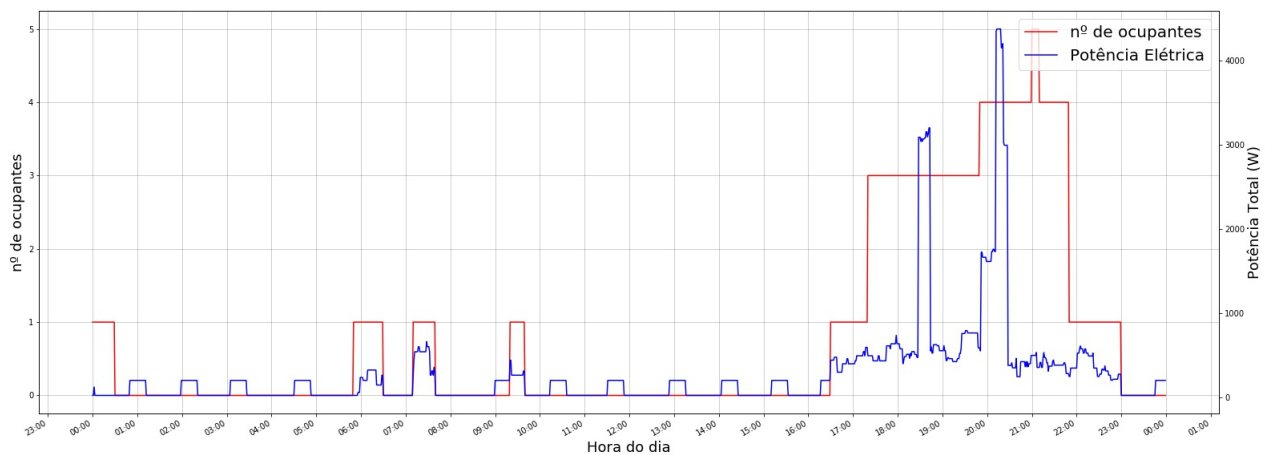


Figura 4.9: Curva de Carga e Ocupação Real

4.5 Cenário 4

O cenário 4 pretende verificar que equipamentos constituem melhores características de entrada dos vários módulos do MDO. Partindo de uma habitação com 15 equipamentos, construiu-se cinco modelos de aprendizagem, treinando-os utilizando diferentes combinações de equipamentos. No sub-cenário 4.1 retiram-se como características do MA o frigorífico e wi-fi tendo piorado o resultado da classificação relativamente aos 69% de exatidão obtidos no sub-cenário 1.1.1. Em seguida no sub-cenário 4.2 retirou-se o consumo de energia relativo à iluminação da habitação, resultando numa pior percentagem de classes corretamente classificadas. No sub-cenário 4.3 apenas foram consideradas como características os diagramas de cargas dos equipamentos com relação direta com a ocupação (e.g. ferro de engomar, aspirador, computador, televisão 1, televisão 2, fogão, forno, micro-ondas, chaleira elétrica, chuveiro elétrico e iluminação) melhorando assim o resultado da classificação face aos sub-cenários 4.1 e 4.2 mas ainda abaixo do resultado obtido para o cenário inicial. O sub-cenário 4.4 permite comprovar a importância do diagrama de carga da iluminação como característica fundamental do MA, deste modo retirando a iluminação e face ao resultado obtido no sub-cenário 4.3 a exatidão do MC desde em média aproximadamente 4%. Por último, o cenário utiliza apenas equipamentos de lazer como as televisões, computador e ainda a iluminação, resultando numa exatidão de classificação de ocupação similar ao sub-cenário 4.3. A Tabela 4.10 apresenta os resultados obtidos em todos os sub-cenários.

Tabela 4.10: Exatidão do MA, MC e MAV utilizando a energia desagregada - Cenário 4

Cenário	nº pessoas	Mês	MA	MC	MAV
4.1	3	Janeiro	89.17%	68.06%	67.15%
4.2			79.67%	64.47%	64.22 %
4.3			87.95%	68.72%	68.31%
4.4			75.80%	64.06%	63.67%
4.5			87.45%	70.03%	69.47%

Na Figura 4.10 está representado o diagrama de caixa da exatidão resultante em cada sub-cenário. Nos sub-cenários 4.1, 4.2 e 4.4 a exatidão de classificação obtida em metade dos dias varia entre 63% e 70%, enquanto que nos sub-cenários 4.3 e 4.5 varia entre 66% e 73%. Em todos os sub-cenários existem valores discrepantes mínimos abaixo dos 50% e, valores discrepantes máximos acima dos 80%. Os modelos implementados nos sub-cenários 4.3 e 4.5 apresentam-se como os mais eficientes, não só por terem o valor mediano mais elevado, mas também por apresentarem valores discrepantes mínimos e máximos mais elevados. Este fato vem comprovar a importância da utilização dos equipamentos de lazer e iluminação como características de entrada do MDO, pois são aqueles diretamente relacionados com a ocupação.

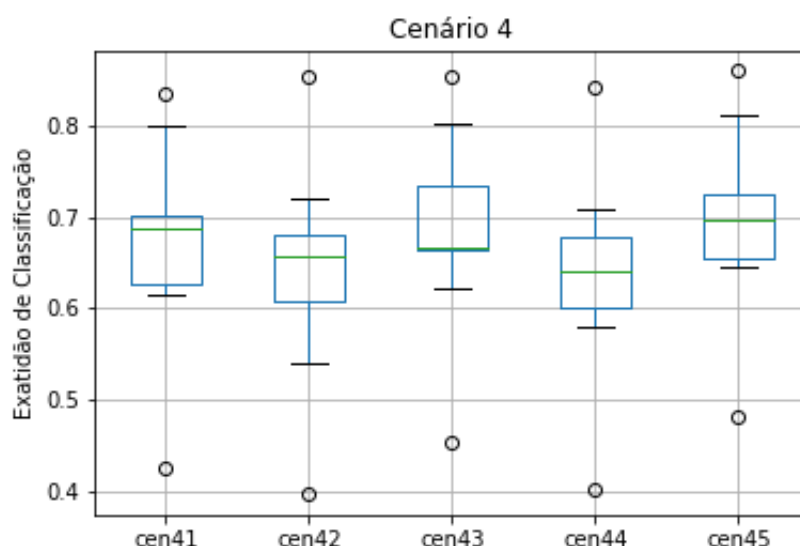


Figura 4.10: Diagrama de caixa referente à exatidão do MDO

Através do modelo implementado no sub-cenário 4.5, determinou-se a ocupação numa habitação durante um determinado dia de Janeiro, utilizando os equipamentos que teoricamente apresentam maior relação com a ocupação.

Assim sendo, recorrendo ao relatório da Tabela 4.11, procedeu-se à análise do MDO na classificação de todas as classes. Como nos casos anteriores, o conjunto de dados não está bem balanceado, pois as classes 0 e 1 são as mais prováveis de ocorrer.

Tabela 4.11: Relatório de Classificação

	Precisão	Recall	F-measure	Suporte
0	98%	99%	98%	980
1	84%	64%	73%	420
2	3%	7%	4%	40
3	0%	0%	0%	0

A exatidão total foi de 86%, enquanto que pela função *F-measure*, a média ponderada entre precisão e *recall* situou-se nos 88%. Individualmente, a classe 0 apresentou o melhor resultado de classificação. Pela matriz de confusão da Tabela 4.12, na classe 0 o número de FP foi apenas 16 e os FN foram 14, resultando numa *F-measure* de 98%. Quanto à classe 1, como o número de FP e FN foi maior a precisão e *recall* foi menor. Por outro lado, em 40 minutos durante os quais a habitação esteve ocupada com 2 pessoas, o MDO apenas considerou 3 minutos, assim o número de FP foi superior aos VP na classificação da classe 2. Por fim, ainda que a habitação nunca estivesse ocupada com 3 pessoas, o MDO, em vez de considerar 1 ocupante, acabou por considerar 3.

Tabela 4.12: Matriz de Confusão

		Classe Prevista			
		0	1	2	3
Classe Actual	0	966	13	1	0
	1	16	268	114	22
	2	0	37	3	0
	3	0	0	0	0

À semelhança dos cenários anteriores, os resultados obtidos através das métricas de avaliação não mostra o comportamento do modelo na classificação em cada classe. Através da Figura 4.11 é possível verificar diferenças entre ocupação real e detetada. Este erro de classificação pode ser identificado pelas oscilações no gráfico da ocupação detetada nos períodos compreendidos entre 6h45-11h e 20h00-23h00.

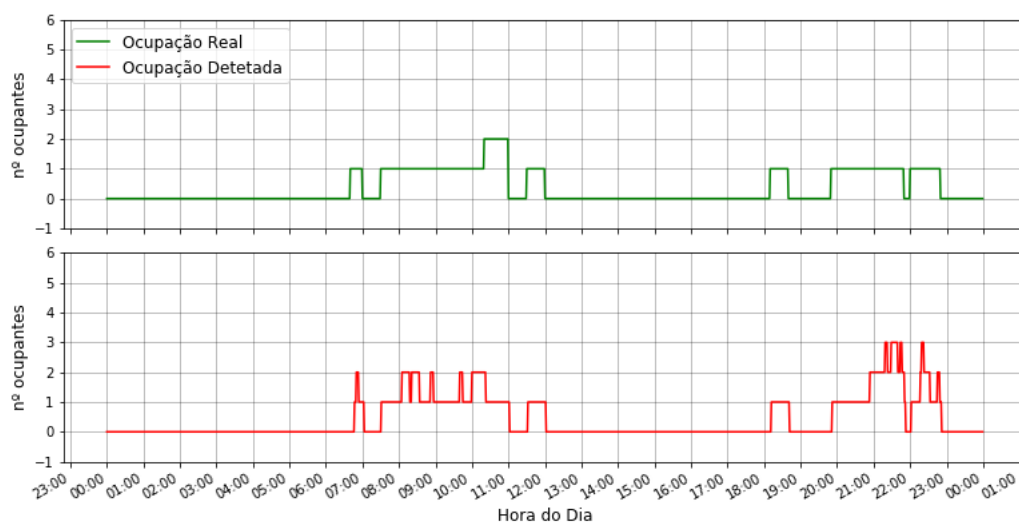


Figura 4.11: Ocupação Real e Detetada - Cenário 4.5

Entre as 06h45 e 07h00 a curva de carga da Figura 4.12 evidencia um aumento do consumo de energia devido ao modo *standby* do computador, televisão e fogão. Contudo, este consumo residual de energia não apresenta um forte indício de ocupação. Durante este período, também os circuitos de iluminação e o ciclo de trabalho do frigorífico contribuíram para o consumo de energia. Com esta informação, o MDO detetou, durante alguns minutos, 2 ocupantes na habitação quando, efetivamente, apenas esteve presente 1 ocupante. Deste modo, conclui-se que o MDO não é sensível a pequenos aumentos do consumo de energia, sendo por isso difícil identificar com rigor o número de ocupantes. Contrariamente, entre as 18h00 e 19h00 surge um grande aumento do consumo de energia devido à utilização do fogão, neste caso o MDO identificou corretamente o número de ocupantes na habitação. Por fim, entre as 20h00 e 23h00, o MDO também apresentou dificuldades no cálculo do número de ocupantes, identificando, com grande frequência, entre 1 e 3 ocupantes, quando na realidade apenas 1 esteve presente. Este erro ocorreu uma vez que o consumo de energia se deveu apenas ao ciclo de trabalho do frigorífico, ao modo *standby* da televisão e fogão, sendo o maior consumo de energia relacionado com a utilização do computador.

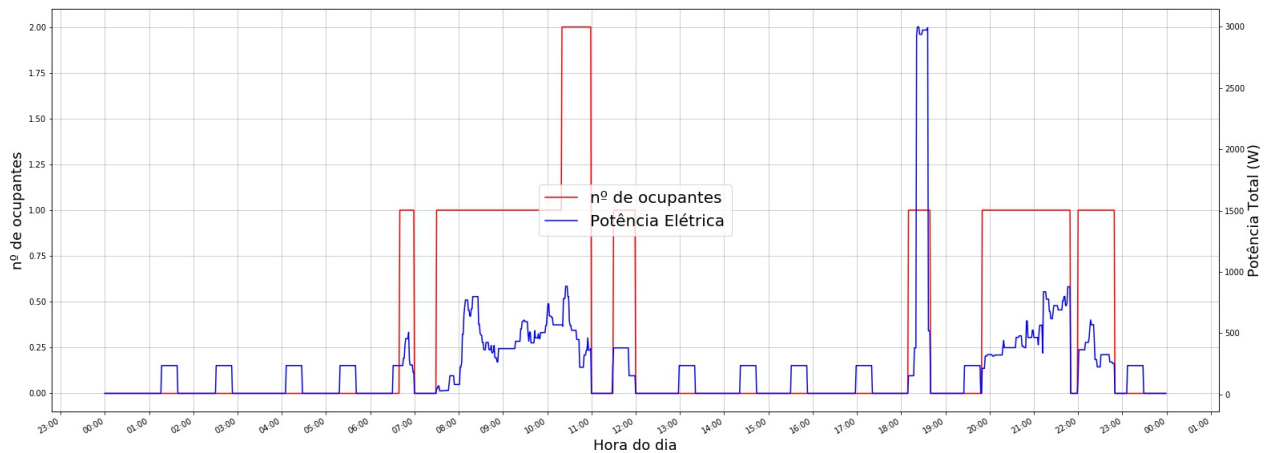


Figura 4.12: Curva de Carga e Ocupação Real

4.6 Análise dos Resultados

Os cenários anteriores permitiram analisar as diversas variáveis que condicionam o processo de detecção de ocupação. A primeira variável prende-se com a construção do modelo de aprendizagem, no qual é necessário escolher o tipo de características necessárias à identificação das novas classes. Foram testados vários casos para verificar se seria mais vantajoso utilizar a energia agregada ou desagregada, concluindo-se, na maioria dos casos, ser melhor utilizar a energia desagregada. Foi também mostrada a importância de usar um modelo de aprendizagem específico para cada mês do ano, ou seja, se pretender determinar o número de ocupantes referente a um determinado mês do ano, os dados usados na fase de treino do algoritmo devem corresponder a esse mês. O número de equipamentos usados como características do modelo de aprendizagem foi uma das variáveis estudadas, tendo-se concluído que um maior número equipamentos não implica, necessariamente, melhor eficácia na detecção de ocupação. Pelos resultados obtidos, mais importante que o número de equipamentos utilizados, está a sua relação com a ocupação. Outra variável importante na detecção do número de ocupantes prende-se com os equipamentos que apresentam um ciclo de trabalho bem definido, como é o caso do frigorífico. Pelos exemplos apresentados anteriormente verifica-se, através das curvas de carga, nomeadamente entre as 12h00 e 18h00, sempre que o frigorífico contribui sozinho para o aumento do consumo de energia, o número de ocupantes na habitação é nulo. Assim sendo, durante este período o MDO sinaliza corretamente a habitação como desocupada. Outra variável importante prende-se com o modo *stand-by* de alguns equipamentos que apesar de contribuírem para o consumo final de energia, podem fornecer informação redundante ao MDO. Assim é importante estabelecer um valor limiar, para eliminar equipamentos com um consumo de energia residual. Por isso, antes de iniciar o processo de construção do modelo de aprendizagem deve-se realizar a seleção das características com melhor relação com a ocupação e que apresentam um consumo de energia significativo.

Segundo os resultados obtidos, conclui-se que devido à simplicidade do algoritmo utilizado e, para um problema onde existe um número significativo de classes, é difícil atingir uma boa performance para todas as classes. Este problema surge, pois, durante o processo de aprendizagem, a base de dados utilizada não apresentava uma distribuição uniforme das classes, sendo sempre mais provável a habitação estar desocupada. Para isso é ser necessário introduzir mais dados na fase de treino do algoritmo.

CONCLUSÃO

Este capítulo apresenta as conclusões gerais do trabalho desenvolvido, tendo em conta os objetivos inicialmente propostos. Posteriormente são apresentadas propostas de trabalhos futuros com vista a melhorar o sistema desenvolvido nesta dissertação.

5.1 Conclusões Finais

A deteção de ocupação em edifícios e habitações é vista como uma importante medida para gerir e reduzir o consumo de energia. Até há relativamente pouco tempo, a deteção de ocupação era realizada maioritariamente através de sensores, aportando, não só, elevados custos de manutenção como também problemas de privacidade. A instalação de contadores inteligentes em grande escala constitui uma alternativa à monitorização de ocupação intrusiva. A partir da informação disponibilizada por estes dispositivos é possível encontrar padrões, na curva de carga que indiquem o número de ocupantes real. A análise de padrões no consumo de energia, relacionados com a interação entre utilizadores e equipamentos elétricos, pode ser realizada através de algoritmos de aprendizagem supervisionada.

O sistema desenvolvido nesta dissertação (i.e. o MDO) tem a capacidade de determinar o número de ocupantes utilizando apenas do consumo de energia. Como tal, na camada de deteção de ocupação, o sistema apresenta um modelo de aprendizagem implementado com recurso ao algoritmo K-nearest neighbor.

Nos testes realizados, o MDO tem a capacidade de detetar o número de ocupantes através da energia agregada e desagregada. De um modo geral, o MDO apresenta melhores resultados quando são utilizadas leituras do consumo de energia de vários equipamentos durante o processo de aprendizagem (i.e. energia desagregada). Por outro lado, o sistema desenvolvido é mais eficiente quanto menos ocupantes a habitação possuir, obtendo, em

média, uma exatidão de 99% num problema de detecção de ocupação binária e 70% numa habitação com três ocupantes. Desde modo, considerando os resultados obtidos, conclui-se que para este sistema ser mais eficaz seria necessário combinar as técnicas de análise do consumo de energia e a utilização de sensores (e.g. CO_2 , indutivo de proximidade de portas e sensores de movimento) ou localização baseada na utilização do sinal wi-fi.

5.2 Trabalhos Futuros

O sistema desenvolvido apresenta várias limitações principalmente relacionadas com o número de ocupantes e com os equipamentos utilizados como características no processo de aprendizagem. Pelos testes realizados ficou evidente que o sistema implementado não reconhece com grande eficácia o número de ocupantes quando vive mais que três pessoas na habitação. A solução para esta limitação passa por utilizar dispositivos auxiliares (e.g. smartphone ou tablet) para registar a entrada e saída de pessoas durante os períodos de maior ocupação (i.e. entre as 08h00-12h00 e 18h-23h00). A partir do histórico de ocupação (e.g. referente à semana anterior) o sistema pode comparar o resultado obtido de modo a reduzir o erro de classificação.

Outra limitação importante do sistema está relacionada com os equipamentos utilizados como características de entrada do modelo de aprendizagem, pois nem sempre os dispositivos escolhidos apresentam um peso significativo na curva de carga da habitação, constituindo informação excessiva e redundante. Deste modo, surge a necessidade de desenvolver um processo para selecionar as características mais importantes, reduzindo assim o número de equipamentos redundantes. Assim, a implementação de algoritmos como o *Principal Component Analysis* ou *Feature Selection Algorithm* constituem uma solução importante na seleção e redução dos equipamentos utilizados como características de entrada do modelo de aprendizagem.

A performance do MDO pode ser melhorada recorrendo a algoritmos de aprendizagem supervisionada mais avançados, como é caso do *Support Vector Machines* utilizado em [10]. Por fim, a implementação de um classificador recorrendo a valores limiares poderia constituir outra solução simples e eficaz. O princípio de execução deste método considera que, a um determinado valor médio do consumo de energia (i.e. obtido durante um dado intervalo de tempo) corresponde um determinado número de ocupantes (e.g. até 100W a habitação está desocupada, entre 100-200W estão 2 ocupantes, etc.).

BIBLIOGRAFIA

- [1] Direção Geral da Geologia e Energia. "Áreas Setoriais, Eficiência Energética: Edifícios". Acedido em 2019. URL: <http://www.dgeg.gov.pt/>.
- [2] Pordata-Base de Dados Portugal Contemporâneo. "Consumo de energia elétrica: total e por tipo de consumo". Acedido em 2019. URL: <https://www.pordata.pt/>.
- [3] Pordata-Base de Dados Portugal Contemporâneo. "Agregados privados com os principais equipamentos elétricos". Acedido em 2019. URL: <https://www.pordata.pt/>.
- [4] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei e T. Weng. "Occupancy-driven Energy Management for Smart Building Automation". Em: *Proceedings of the 2Nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*. 2010.
- [5] J. Lo e A. Novoselac. "Localized air-conditioning with occupancy control in an open office". Em: *Energy and Buildings - ENERG BLDG* 42 (2010).
- [6] Y. Zhu, T. Batten, H. L. Noboa, D. E. Claridge, W. d. Turner, M. Liu, J. Zhou, C. Cameron, D. R. Keeble e R. Hirschak. "Optimization Control Strategies for HVAC Terminal Boxes". Em: 2000.
- [7] D. Chen, S. Barker, A. Subbaswamy, D. Irwin e P. Shenoy. "Non-Intrusive Occupancy Monitoring Using Smart Meters". Em: *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. 2013.
- [8] Parlamento Europeu. "Directiva 2009/72/CE do Parlamento Europeu e do Conselho de 13 de Julho de 2009". Acedido em 2019. URL: <https://eur-lex.europa.eu/legal-content/pt/TXT/?uri=CELEX%3A32009L0072>.
- [9] ERSE-Entidade Reguladora dos Serviços Energéticos. "Contadores Inteligentes de Eletricidade - Estudo Previsto na Portaria n.º 231/2013". Em: Acedido em 2019. URL: <https://www.erse.pt/search/?query=contadores+inteligentes>.
- [10] W. Kleiminger, H.-J. Appellrath, A. Dey, S. Santini, Kurzfassung, A. Bernauer, B. Ostermeier, R. Adelmann, M. Bâce, M. George, V. Trifa, D. Guinard, C. Floerkemeier, M. Weiss, E. Schaper, D. Pauli, A. Droescher, S. Kilcher, A. Brauchli, M. Spiegel e C. Stüclberger. "Occupancy Sensing and Prediction for Automated Energy Savings". Em: 2015.

- [11] V. L. Erickson, Y. Lin, A. Kamthe, R. Brahme, A. Surana, A. E. Cerpa, M. D. Sohn e S. Narayanan. "Energy Efficient Building Environment Control Strategies Using Real-time Occupancy Measurements". Em: *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*. 2009.
- [12] EDP Distribuição. "Energy Box". Acedido em 2019. URL: <https://www.edpdistribuicao.pt/pt-pt/contadores-inteligentes>.
- [13] S. Wang e X. Jin. "CO₂ Based Occupancy Detection for On-Line Outdoor Air Flow Control". Em: *Indoor and Built Environment - INDOOR BUILT ENVIRON* (1998).
- [14] J. Lu, T. I. Sookoor, V. Srinivasan, G. Gao, B. Holben, J. A. Stankovic, E. Field e K. Whitehouse. "The smart thermostat: using occupancy sensors to save energy in homes". Em: 2010.
- [15] T. Hnat, V. Srinivasan, J. Lu, T. Sookoor, R. Dawson, J. A. Stankovic e K. Whitehouse. "The hitchhiker's guide to successful residential sensing deployments". Em: 2011.
- [16] W. Kleiminger, C. Beckel, T. Staake e S. Santini. "Occupancy Detection from Electricity Consumption Data". Em: *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. BuildSys'13. 2013.
- [17] W. Kleiminger, C. Beckel e S. Santini. "Household Occupancy Monitoring Using Electricity Meters". Em: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '15. 2015.
- [18] M. Jin, R. Jia e C. J. Spanos. "Virtual Occupancy Sensing: Using Smart Meters to Indicate Your Presence". Em: *IEEE Transactions on Mobile Computing* (2017).
- [19] S. Tarzia, R. Dick, P. Dinda e G. Memik. "Sonar-based measurement of user presence and attention". Em: 2009.
- [20] R. Dodier, G. Henze, D. Tiller e X. Guo. "Building occupancy detection through sensor belief networks". Em: *Energy and Buildings* (2006).
- [21] E. Hailemariam, R. Goldstein, R. Attar e A. Khan. "Real-time Occupancy Detection Using Decision Trees with Multiple Sensor Types". Em: *Proceedings of the 2011 Symposium on Simulation for Architecture and Urban Design*. 2011.
- [22] R. M. A. Lopes. "Monitorização não Intrusiva de Cargas". Mestrado em Engenharia Eletrotécnica e de Computadores. Lisboa: Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, 2012.
- [23] J. F. S. Gomes. "Sistema de Monitorização não Intrusiva de Cargas". Mestrado em Engenharia Eletrotécnica e de Computadores. Lisboa: Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, 2017.
- [24] G. W. Hart. "Nonintrusive appliance load monitoring". Em: *Proceedings of the IEEE* (1992).

-
- [25] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake e S. Santini. "The ECO Data Set and the Performance of Non-Intrusive Load Monitoring Algorithms". Em: nov. de 2014.
 - [26] A. W. Whitney. "A Direct Method of Nonparametric Measurement Selection". Em: *IEEE Trans. Comput.* (1971).
 - [27] T. Vafeiadis, S. Zikos, G. Stavropoulos, D. Ioannidis, S. Krinidis, D. Tzovaras e K. Moustakas. "Machine Learning Based Occupancy Detection via the Use of Smart Meters". Em: 2017.
 - [28] G. Tang, K. Wu, J. Lei e W. Xiao. "The meter tells you are at home! Non-intrusive occupancy detection via load curve data". Em: *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. 2015.
 - [29] A. Akbar, M. Nati, F. Carrez e K. Moessner. "Contextual Occupancy Detection for Smart Office by Pattern Recognition of Electricity Consumption Data". Em: 2015.
 - [30] Scikit Learn. "*Nearest Neighbor*". 2018. URL: <https://scikit-learn.org/stable/modules/neighbors.html>.
 - [31] Scikit Learn. "*API Reference*". 2018. URL: <https://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics>.
 - [32] Jupyter Notebook. "*Some information about the Jupyter Project and Community*". Acedido em 2019. URL: <https://jupyter.org/about>.
 - [33] Anaconda. "*Solutions for Data Science Practitioners and Enterprise Machine Learning*". Acedido em 2019. URL: <https://www.anaconda.com/>.
 - [34] I. Richardson, M. Thomson, D. Infield e C. Clifford. "Domestic electricity use: a high-resolution energy demand model". Em: (2010). URL: https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/6997?mode=simple&submit_simple=Show+simple+item+record.
 - [35] Pandas. "*Python Data Analysis Library*". Acedido em 2019. URL: <https://pandas.pydata.org/>.

